ISO 25964-1: a new standard for development of thesauri and exchange of thesaurus data

Stella G Dextre Clarke and

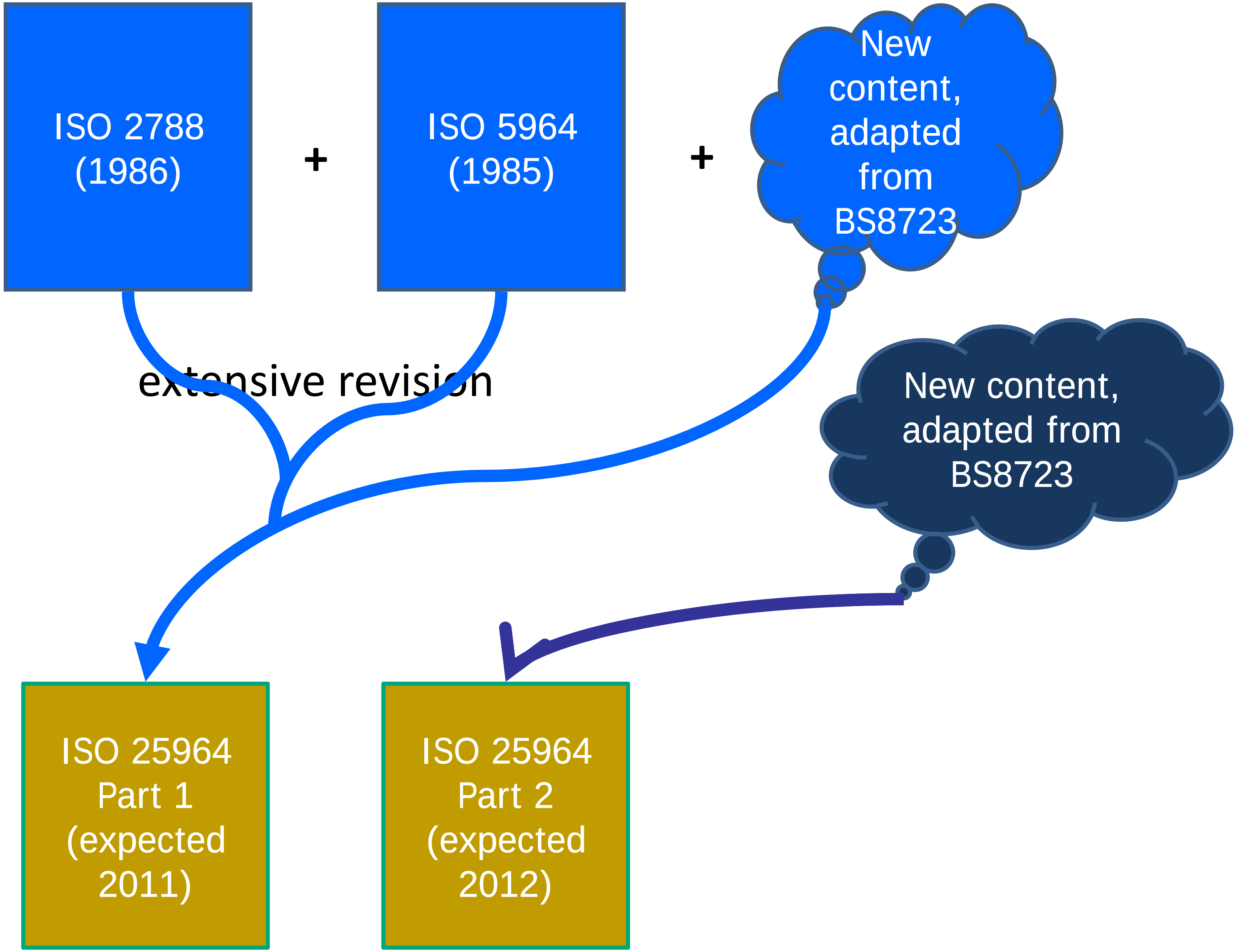
Johan De Smedt

What is ISO 25964?

ISO 25964: Thesauri and interoperability with other vocabularies

  - Part 1: Thesauri for information retrieval
  - Part 2: Interoperability with other vocabularies

- It updates ISO 2788 and ISO 5964
- based on BS 8723, with much reworking
- Part 1, **published in August 2011**, covers monolingual and multilingual thesauri
- Part 2, to be published in 2012, covers mapping between thesauri and other types of vocabulary
- information retrieval seen as main application; mapping applies to index terms or to search terms



ISO 2788 (1986)

+

ISO 5964 (1985)

+

New content, adapted from BS8723

extensive revision

New content, adapted from BS8723

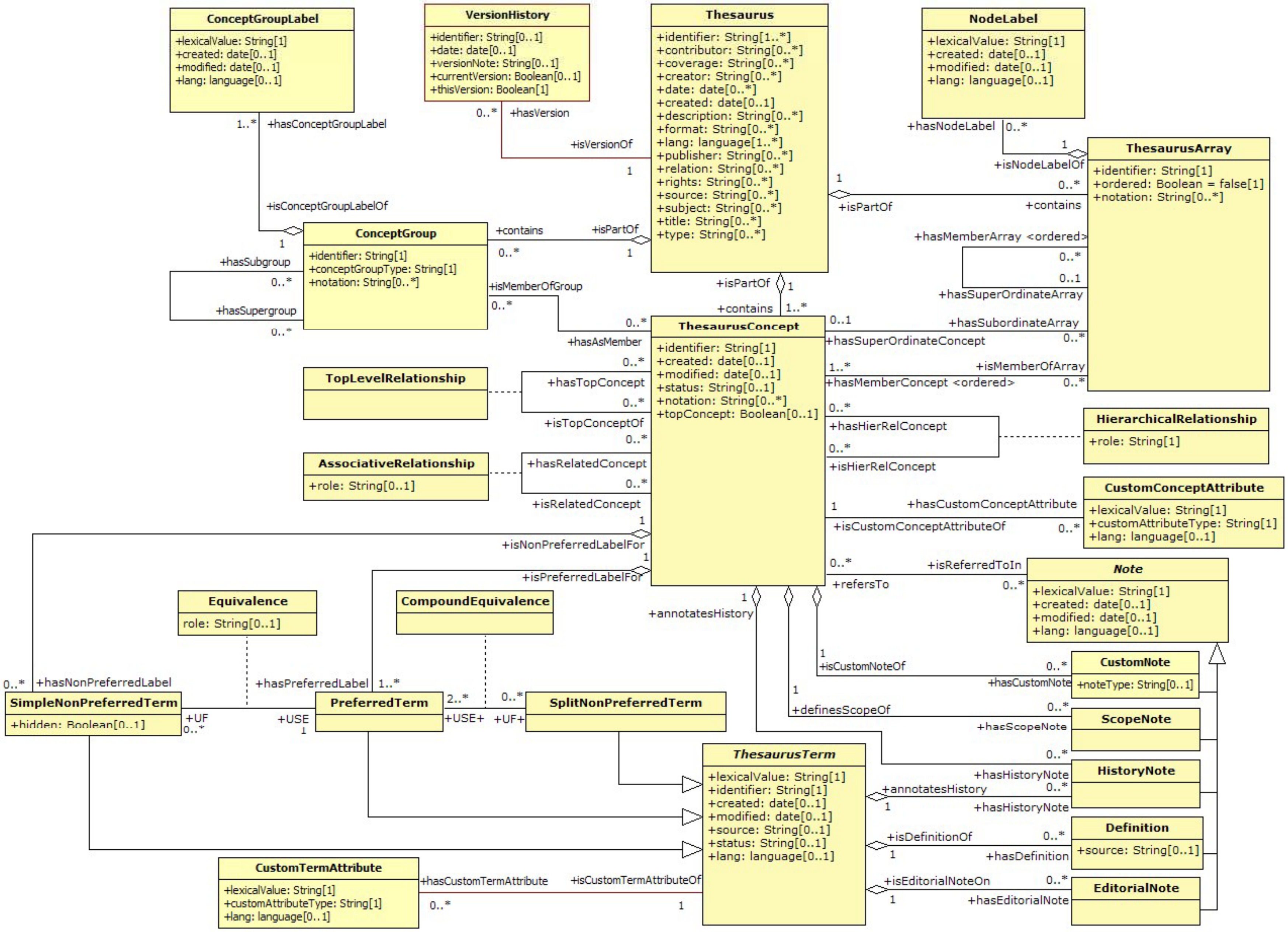
ISO 25964 Part 1 (expected 2011)

ISO 25964 Part 2 (expected 2012)

What’s in Part 1?

All that was in ISO 2788 and ISO 5964, revised and extended to include:

- thesaurus content and construction, mono- or multi-lingual.
- guidance on applying facet analysis to thesauri
- guidance on managing thesaurus development and maintenance
- functional requirements for software to manage thesauri
- a data model and derived XML schema


ConceptGroupLabel
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]
VersionHistory
+identifier: String[0..1]
+date: date[0..1]
+versionNote: String[0..1]
+currentVersion: Boolean[0..1]
+thisVersion: Boolean[1]
Thesaurus
+identifier: String[1..*]
+contributor: String[0..*]
+coverage: String[0..*]
+creator: String[0..*]
+date: date[0..*]
+created: date[0..1]
+description: String[0..*]
+format: String[0..*]
+lang: language[1..*]
+publisher: String[0..*]
+relation: String[0..*]
+rights: String[0..*]
+source: String[0..*]
+subject: String[0..*]
+title: String[0..*]
+type: String[0..*]
NodeLabel
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]
ThesaurusArray
+identifier: String[1]
+ordered: Boolean = false[1]
+notation: String[0..*]
ConceptGroup
+identifier: String[1]
+conceptGroupType: String[1]
+notation: String[0..*]
ThesaurusConcept
+identifier: String[1]
+created: date[0..1]
+modified: date[0..1]
+status: String[0..1]
+notation: String[0..*]
+topConcept: Boolean[0..1]
TopLevelRelationship
AssociativeRelationship
+role: String[0..1]
HierarchicalRelationship
+role: String[1]
CustomConceptAttribute
+lexicalValue: String[1]
+customAttributeType: String[1]
+lang: language[0..1]
Note
+lexicalValue: String[1]
+created: date[0..1]
+modified: date[0..1]
+lang: language[0..1]
CustomNote
+noteType: String[0..1]
ScopeNote
HistoryNote
Definition
+source: String[0..1]
EditorialNote
Equivalence
role: String[0..1]
CompoundEquivalence
SimpleNonPreferredTerm
+hidden: Boolean[0..1]
PreferredTerm
SplitNonPreferredTerm
ThesaurusTerm
+lexicalValue: String[1]
+identifier: String[1]
+created: date[0..1]
+modified: date[0..1]
+source: String[0..1]
+status: String[0..1]
+lang: language[0..1]
CustomTermAttribute
+lexicalValue: String[1]
+customAttributeType: String[1]
+lang: language[0..1]
1..*
+hasConceptGroupLabel
+isConceptGroupLabelOf
1
0..*
+hasVersion
+isVersionOf
1
+hasNodeLabel
0..*
1
+isNodeLabelOf
1
0..*
+isPartOf
+contains
+hasMemberArray <ordered>
0..*
0..1
+hasSuperOrdinateArray
+contains
+isPartOf
0..*
1
+hasSubgroup
0..*
+hasSupergroup
0..*
+isMemberOfGroup
0..*
+isPartOf
1
+contains
1..*
0..*
+hasAsMember
0..1
+hasSubordinateArray
+hasSuperOrdinateConcept
0..*
1..*
+isMemberOfArray
+hasMemberConcept <ordered>
0..*
0..*
+hasTopConcept
0..*
+isTopConceptOf
0..*
+hasHierRelConcept
0..*
+isHierRelConcept
0..*
+hasRelatedConcept
0..*
+isRelatedConcept
1
+hasCustomConceptAttribute
+isCustomConceptAttributeOf
0..*
1
+isNonPreferredLabelFor
1
+isPreferredLabelFor
0..*
+isReferredToIn
+refersTo
0..*
1
+annotatesHistory
1
+isCustomNoteOf
0..*
+hasCustomNote
1
+definesScopeOf
0..*
+hasScopeNote
0..*
+hasHistoryNote
+annotatesHistory
0..*
+hasHistoryNote
1
+isDefinitionOf
0..*
1
+hasDefinition
+isEditorialNoteOn
0..*
1
+hasEditorialNote
0..*
+hasNonPreferredLabel
+hasPreferredLabel
1..*
2..*
0..*
+UF
0..*
+USE
1
+USE+
+UF+
+hasCustomTermAttribute
+isCustomTermAttributeOf
0..*
1

What’s in Part 2?

- Models for mapping
- Guidelines for mapping
  - Recommendations on mapping types
  - How to handle pre-coordination
  - Mapping to vocabularies other than thesauri: classification schemes, file plans, taxonomies, subject heading schemes, ontologies, synonym rings, terminologies and name authority lists
- Brief guidance on handling mappings data

Want a copy of ISO 25964-1 ?

- Download it from ISO at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53657
- Order it from your national standards body (e.g. BSI, DIN, ANSI, AFNOR)
- Some public/academic reference libraries may stock it
- It is not cheap to purchase
- However, the XML schema for exchange of thesaurus data is in an Annex which is available online without charge or password control. Go to http://www.niso.org/schemas/iso25964/

Want a copy of ISO 25964-2 ?

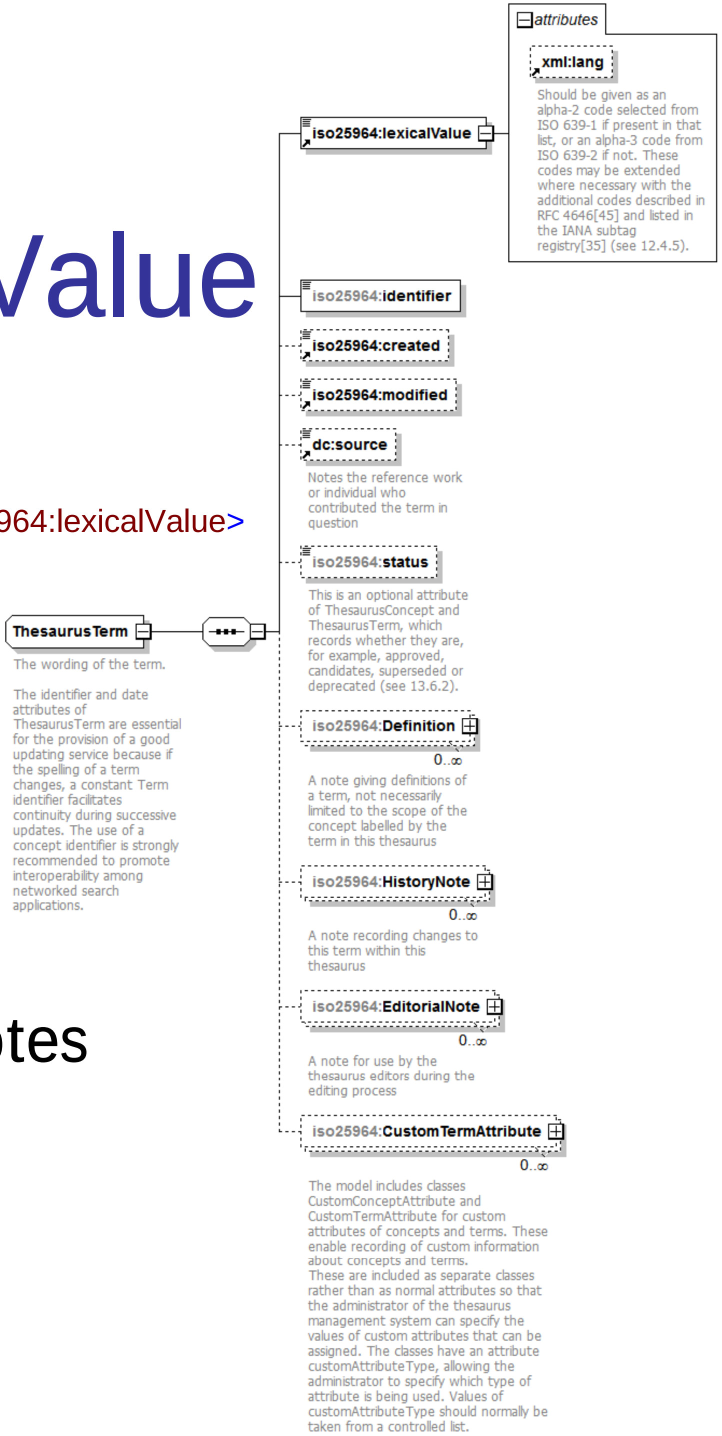
- A draft will be issued later in 2011, “ISO DIS 25964-2”, with the hope of attracting comments from potential users
- The official way to get it is through your national standards body (e.g. BSI, DIN)
- Distribution policies vary from one country to another; but for a couple of months the draft should be available online free of charge and free of passwords, on the BSI site.
- Send me an email and I’ll alert you when the DIS is released. stella@lukehouse.org

The XML-Schema

- Based on the UML model to capture a maximum of the specifications of the standard
- Use of Dublin Core elements

Terms and lexicalValue

- Example
  - <iso25964:lexicalValue xml:lang="en">clothing</iso25964:lexicalValue>
- Exactly one lexical value
  - with optional language required in multi-lingual thesaurus
- A required identifier
- Optional
  - dates, source reference, notes
  - custom attributes
    - name-type, value



Should be given as an alpha-2 code selected from ISO 639-1 if present in that list, or an alpha-3 code from ISO 639-2 if not. These codes may be extended where necessary with the additional codes described in RFC 4646[45] and listed in the IANA subtag registry[35] (see 12.4.5).

Notes the reference work or individual who contributed the term in question

This is an optional attribute of ThesaurusConcept and ThesaurusTerm, which records whether they are, for example, approved, candidates, superseded or deprecated (see 13.6.2).

The wording of the term.

The identifier and date attributes of ThesaurusTerm are essential for the provision of a good updating service because if the spelling of a term changes, a constant Term identifier facilitates continuity during successive updates. The use of a concept identifier is strongly recommended to promote interoperability among networked search applications.

A note giving definitions of a term, not necessarily limited to the scope of the concept labelled by the term in this thesaurus

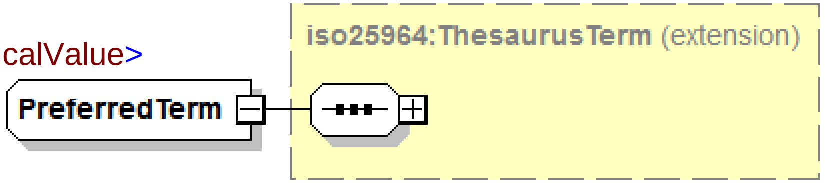
A note recording changes to this term within this thesaurus

A note for use by the thesaurus editors during the editing process

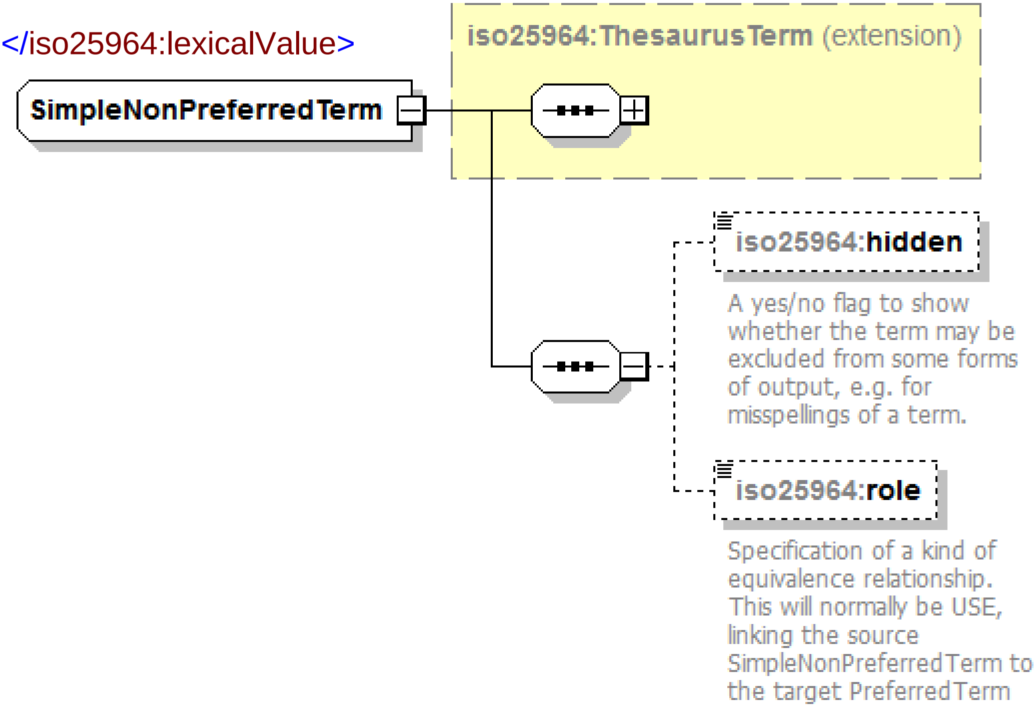
The model includes classes CustomConceptAttribute and CustomTermAttribute for custom attributes of concepts and terms. These enable recording of custom information about concepts and terms.
These are included as separate classes rather than as normal attributes so that the administrator of the thesaurus management system can specify the values of custom attributes that can be assigned. The classes have an attribute customAttributeType, allowing the administrator to specify which type of attribute is being used. Values of customAttributeType should normally be taken from a controlled list.

Preferred Term
Simple non-Preferred Term

- <iso25964:PreferredTerm>
-    <iso25964:lexicalValue xml:lang="en">abattoirs</iso25964:lexicalValue>
-    <iso25964:identifier>L11</iso25964:identifier>
- </iso25964:PreferredTerm>

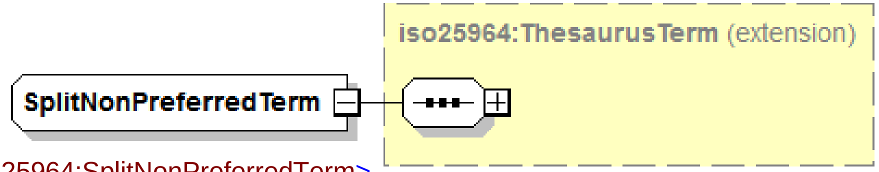


- <iso25964:SimpleNonPreferredTerm>
-    <iso25964:lexicalValue xml:lang="en">abatoirs</iso25964:lexicalValue>
-    <iso25964:identifier>L12</iso25964:identifier>
-    <iso25964:hidden>true</iso25964:hidden>
-    <iso25964:role>MS</iso25964:role>
- </iso25964:SimpleNonPreferredTerm>

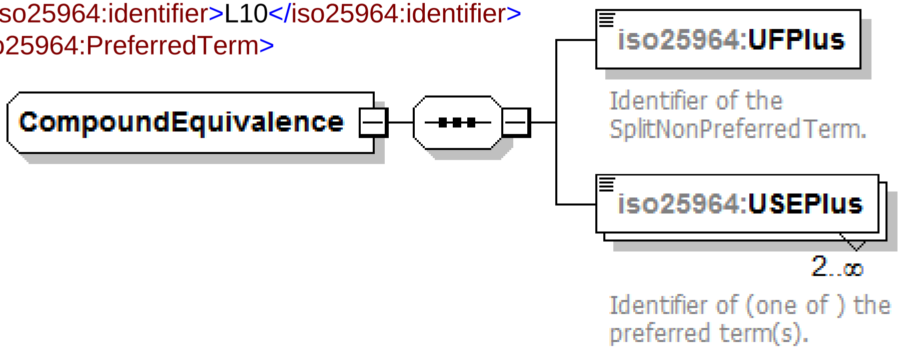

iso25964:ThesaurusTerm (extension)
SimpleNonPreferredTerm
iso25964:hidden
A yes/no flag to show whether the term may be excluded from some forms of output, e.g. for misspellings of a term.
iso25964:role
Specification of a kind of equivalence relationship. This will normally be USE, linking the source SimpleNonPreferredTerm to the target PreferredTerm

Split non-preferred Term Compound equivalence

- Split non preferred term
  - one (term) identifier
  - one lexical value
  - May engage in one or more compound equivalence relationships
- Compound Equivalence:
  - UFPlus
    - identifier of Split non Preferred Term
  - USEPlus
    - identifier of Preferred Term
    - at least 2



```
<iso25964:SplitNonPreferredTerm>
   <iso25964:lexicalValue xml:lang="en">coal mining</iso25964:lexicalValue>
   <iso25964:identifier>SL1</iso25964:identifier>
</iso25964:SplitNonPreferredTerm>
<iso25964:CompoundEquivalence>
   <iso25964:UFPlus>SL1</iso25964:UFPlus>
   <iso25964:USEPlus>L9</iso25964:USEPlus>
   <iso25964:USEPlus>L10</iso25964:USEPlus>
</iso25964:CompoundEquivalence>

<iso25964:PreferredTerm>
   <iso25964:lexicalValue xml:lang="en">coal</iso25964:lexicalValue>
   <iso25964:identifier>L9</iso25964:identifier>
</iso25964:PreferredTerm>

<iso25964:PreferredTerm>
   <iso25964:lexicalValue xml:lang="en">mining</iso25964:lexicalValue>
   <iso25964:identifier>L10</iso25964:identifier>
</iso25964:PreferredTerm>
```


CompoundEquivalence
iso25964:UFPlus
Identifier of the SplitNonPreferredTerm.
iso25964:USEPlus
2..∞
Identifier of (one of) the preferred term(s).

Concept and Equivalent Terms (1/2)

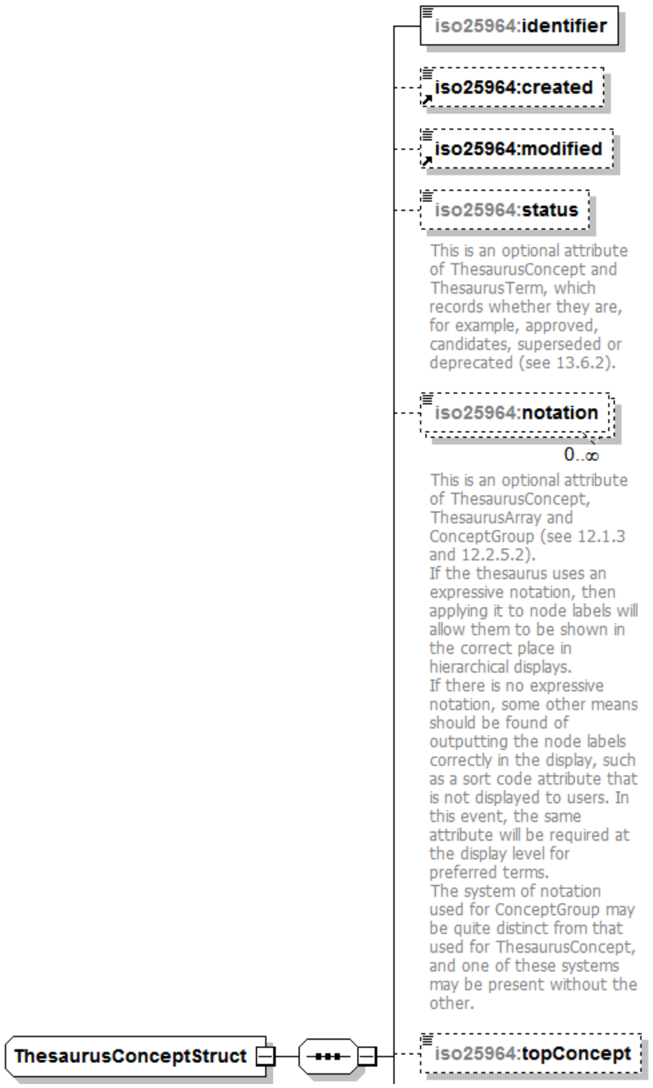
Equivalent preferred and non-preferred terms

```
<iso25964:ThesaurusConcept>
  <iso25964:identifier>C11</iso25964:identifier>
  <iso25964:PreferredTerm>
    <iso25964:lexicalValue xml:lang="en">abattoirs</iso25964:lexicalValue>
    <iso25964:identifier>L11</iso25964:identifier>
  </iso25964:PreferredTerm>
  <iso25964:SimpleNonPreferredTerm>
    <iso25964:lexicalValue xml:lang="en">abatoirs</iso25964:lexicalValue>
    <iso25964:identifier>L12</iso25964:identifier>
    <iso25964:hidden>true</iso25964:hidden>
    <iso25964:role>MS</iso25964:role>
  </iso25964:SimpleNonPreferredTerm>
</iso25964:ThesaurusConcept>
```

Equivalent preferred terms in a different language

```
<iso25964:ThesaurusConcept>
  <iso25964:identifier>C10</iso25964:identifier>
  <iso25964:PreferredTerm>
    <iso25964:lexicalValue xml:lang="en">mining</iso25964:lexicalValue>
    <iso25964:identifier>L10</iso25964:identifier>
  </iso25964:PreferredTerm>
  <iso25964:PreferredTerm>
    <iso25964:lexicalValue xml:lang="fr">exploitation
                                  minière</iso25964:lexicalValue>
    <iso25964:identifier>L10.fr</iso25964:identifier>
  </iso25964:PreferredTerm>
</iso25964:ThesaurusConcept>
```

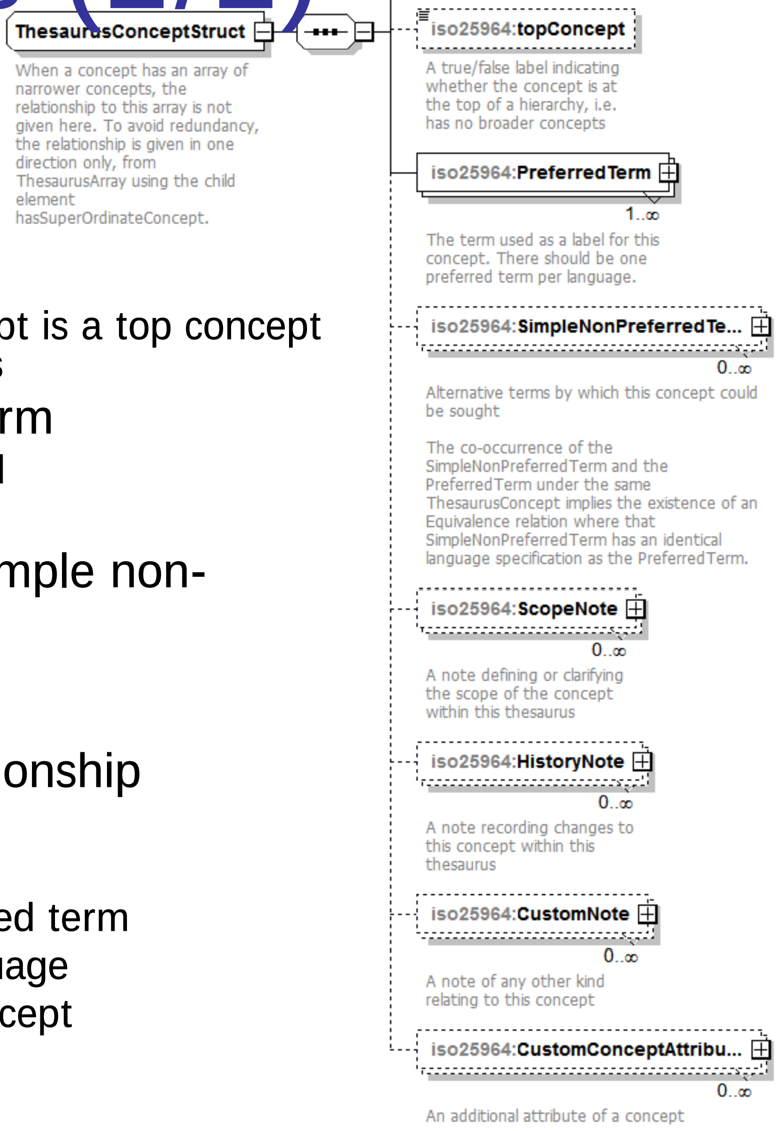
Concept and Equivalent Terms (2/2)



This is an optional attribute of ThesaurusConcept and ThesaurusTerm, which records whether they are, for example, approved, candidates, superseded or deprecated (see 13.6.2).

This is an optional attribute of ThesaurusConcept, ThesaurusArray and ConceptGroup (see 12.1.3 and 12.2.5.2).
If the thesaurus uses an expressive notation, then applying it to node labels will allow them to be shown in the correct place in hierarchical displays.
If there is no expressive notation, some other means should be found of outputting the node labels correctly in the display, such as a sort code attribute that is not displayed to users. In this event, the same attribute will be required at the display level for preferred terms.
The system of notation used for ConceptGroup may be quite distinct from that used for ThesaurusConcept, and one of these systems may be present without the other.

- Top Concept
  - True: the concept is a top concept of the thesaurus
- One Preferred Term
  - per Concept and
  - per language
- Any number of simple non-preferred term
  - per Concept
  - with language
- Equivalence relationship between
  - preferred term
  - any non-preferred term
  - with same language
  - under same concept



When a concept has an array of narrower concepts, the relationship to this array is not given here. To avoid redundancy, the relationship is given in one direction only, from ThesaurusArray using the child element hasSuperOrdinateConcept.

A true/false label indicating whether the concept is at the top of a hierarchy, i.e. has no broader concepts

The term used as a label for this concept. There should be one preferred term per language.

Alternative terms by which this concept could be sought

The co-occurrence of the SimpleNonPreferredTerm and the PreferredTerm under the same ThesaurusConcept implies the existence of an Equivalence relation where that SimpleNonPreferredTerm has an identical language specification as the PreferredTerm.

A note defining or clarifying the scope of the concept within this thesaurus

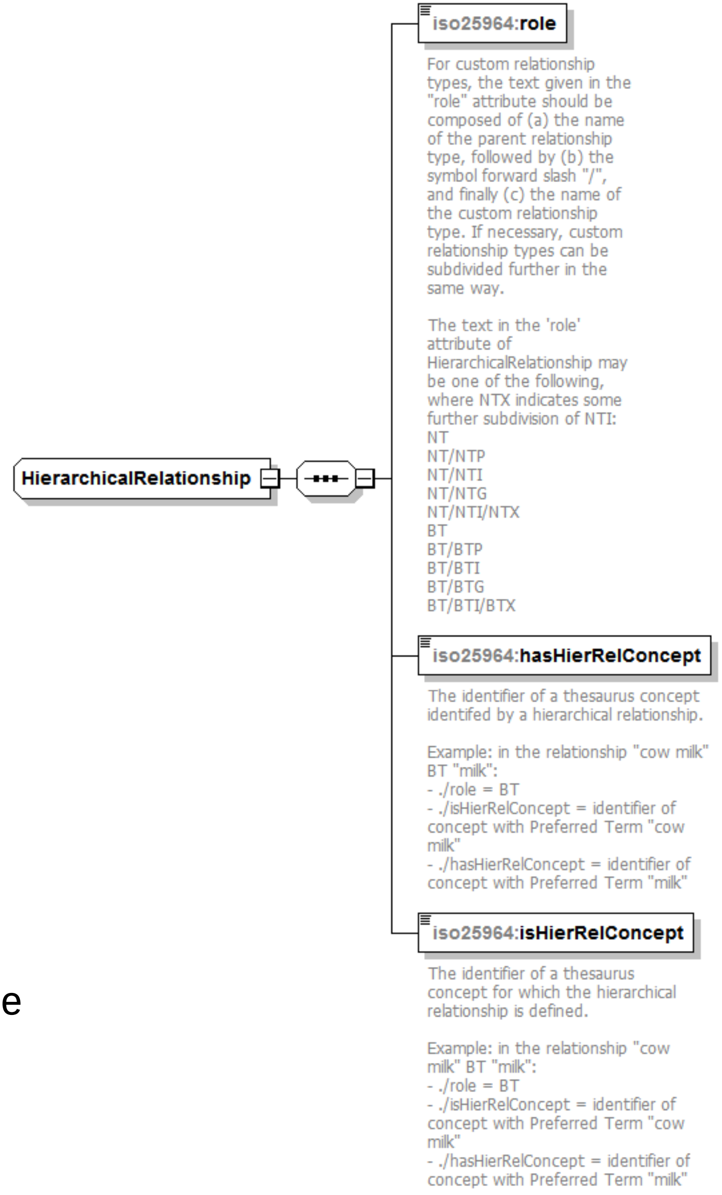
A note recording changes to this concept within this thesaurus

A note of any other kind relating to this concept

An additional attribute of a concept

Concept relations Hierarchy (1/3)

- Milk
  - Cow milk
- “Cow milk” BT “Milk”
- Coded as
  - role: BT
  - isHierRelConcept
    - references a concept identifier
    - is (subject): = identifier of “Cow milk”
  - hasHierRelConcept
    - references a concept identifier
    - has (object): = identifier of “milk”
- Note:
  - For any given concept, there can be more than one top concept



For custom relationship types, the text given in the "role" attribute should be composed of (a) the name of the parent relationship type, followed by (b) the symbol forward slash "/", and finally (c) the name of the custom relationship type. If necessary, custom relationship types can be subdivided further in the same way.

The text in the 'role' attribute of HierarchicalRelationship may be one of the following, where NTX indicates some further subdivision of NTI:
NT
NT/NTP
NT/NTI
NT/NTG
NT/NTI/NTX
BT
BT/BTP
BT/BTI
BT/BTG
BT/BTI/BTX

The identifier of a thesaurus concept identifed by a hierarchical relationship.

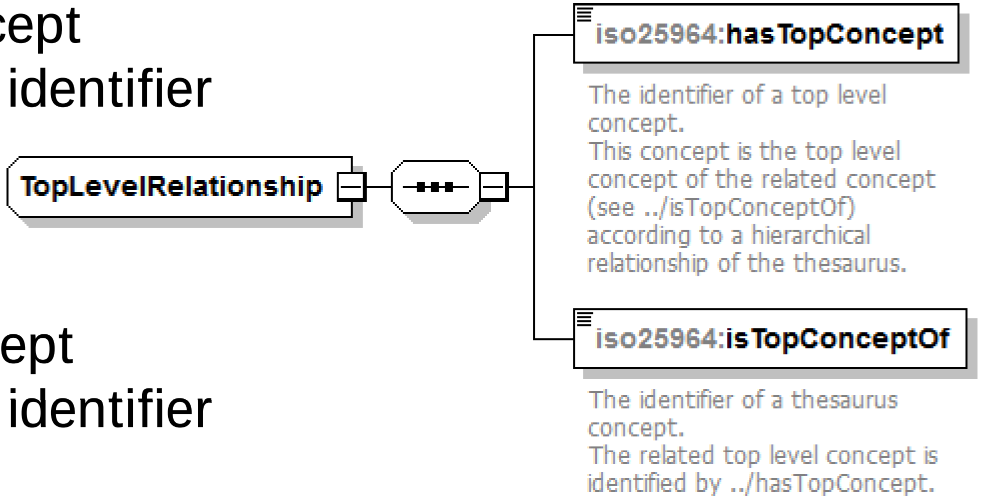
Example: in the relationship "cow milk" BT "milk":
- ./role = BT
- ./isHierRelConcept = identifier of concept with Preferred Term "cow milk"
- ./hasHierRelConcept = identifier of concept with Preferred Term "milk"

The identifier of a thesaurus concept for which the hierarchical relationship is defined.

Example: in the relationship "cow milk" BT "milk":
- ./role = BT
- ./isHierRelConcept = identifier of concept with Preferred Term "cow milk"
- ./hasHierRelConcept = identifier of concept with Preferred Term "milk"

Concept relations Top-Level (2/3)

- isTopConceptOf
  - is (subject): any concept
  - references a concept identifier
- hasTopConcept
  - has (object):
    The related Top concept
  - references a concept identifier
- Note:
  - For any given concept, there can be more than one top concept



The identifier of a top level concept.
This concept is the top level concept of the related concept (see ../isTopConceptOf) according to a hierarchical relationship of the thesaurus.

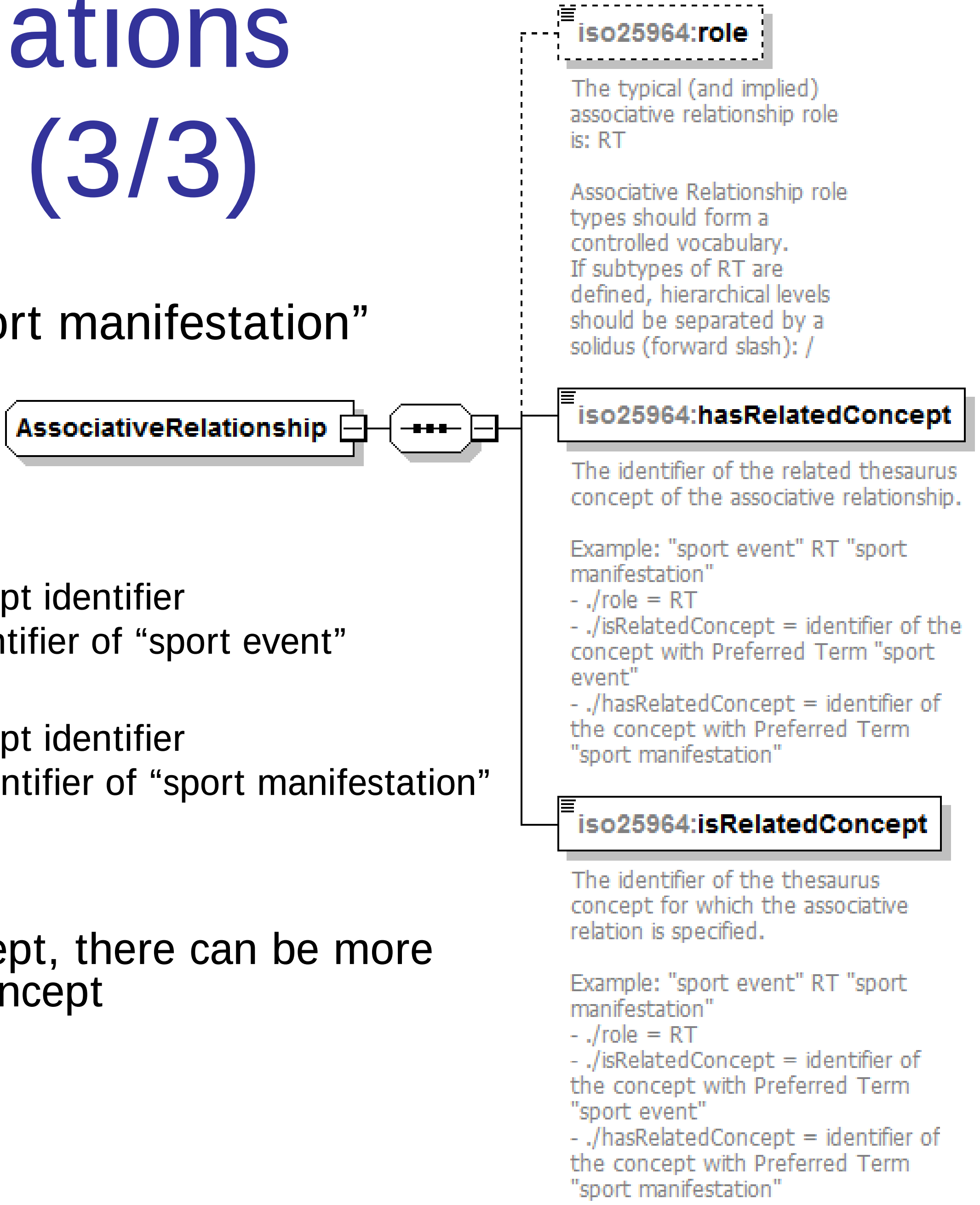
The identifier of a thesaurus concept.
The related top level concept is identified by ../hasTopConcept.

Concept relations Associative (3/3)

- “sport event” RT “sport manifestation”

- Coded as
  - role: RT
  - isRelatedConcept
    - references a concept identifier
    - is (subject): = identifier of “sport event”
  - hasRelatedConcept
    - references a concept identifier
    - has (object): = identifier of “sport manifestation”

- Note:
  - For any given concept, there can be more than one related concept



The typical (and implied) associative relationship role is: RT

Associative Relationship role types should form a controlled vocabulary.
If subtypes of RT are defined, hierarchical levels should be separated by a solidus (forward slash): /

The identifier of the related thesaurus concept of the associative relationship.

Example: "sport event" RT "sport manifestation"
- ./role = RT
- ./isRelatedConcept = identifier of the concept with Preferred Term "sport event"
- ./hasRelatedConcept = identifier of the concept with Preferred Term "sport manifestation"

The identifier of the thesaurus concept for which the associative relation is specified.

Example: "sport event" RT "sport manifestation"
- ./role = RT
- ./isRelatedConcept = identifier of the concept with Preferred Term "sport event"
- ./hasRelatedConcept = identifier of the concept with Preferred Term "sport manifestation"

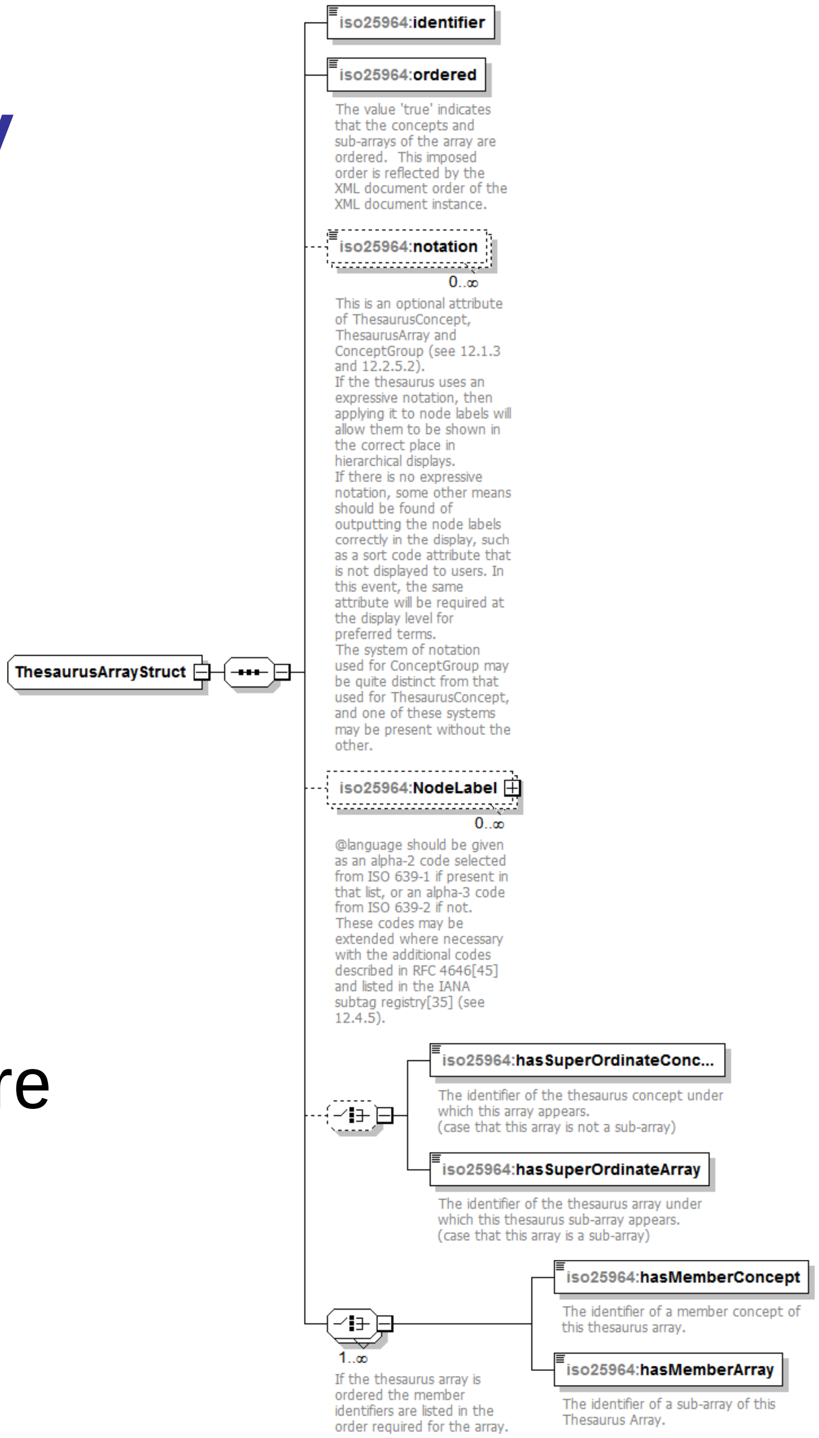
Thesaurus Array (1/2)

```
<iso25964:ThesaurusArray>
   <iso25964:identifier>A1</iso25964:identifier>
   <iso25964:ordered>true</iso25964:ordered>
   <iso25964:NodeLabel>
      <iso25964:lexicalValue xml:lang="en">age group</iso25964:lexicalValue>
   </iso25964:NodeLabel>
   <iso25964:hasSuperOrdinateConcept>C5</iso25964:hasSuperOrdinateConcept>
   <iso25964:hasMemberConcept>C6</iso25964:hasMemberConcept>
   <iso25964:hasMemberConcept>C7</iso25964:hasMemberConcept>
   <iso25964:hasMemberConcept>C8</iso25964:hasMemberConcept>
</iso25964:ThesaurusArray>
```

```
<iso25964:ThesaurusConcept>
  <iso25964:identifier>C5</iso25964:identifier>
  <iso25964:PreferredTerm>
    <iso25964:lexicalValue
                      xml:lang="en">people</iso25964:lexicalValue>
    <iso25964:identifier>L5</iso25964:identifier>
    </iso25964:PreferredTerm>
</iso25964:ThesaurusConcept>

<iso25964:ThesaurusConcept>
  <iso25964:identifier>C6</iso25964:identifier>
   <iso25964:PreferredTerm>
      <iso25964:lexicalValue
                      xml:lang="en">children</iso25964:lexicalValue>
  <iso25964:identifier>L6</iso25964:identifier>
  </iso25964:PreferredTerm>
</iso25964:ThesaurusConcept>

<iso25964:ThesaurusConcept>
  <iso25964:identifier>C7</iso25964:identifier>
  <iso25964:PreferredTerm>
    <iso25964:lexicalValue xml:lang="en">youths</iso25964:lexicalValue>
    <iso25964:identifier>L7</iso25964:identifier>
  </iso25964:PreferredTerm>
</iso25964:ThesaurusConcept>

<iso25964:ThesaurusConcept>
  <iso25964:identifier>C8</iso25964:identifier>
    <iso25964:PreferredTerm>
    <iso25964:lexicalValue xml:lang="en">adults</iso25964:lexicalValue>
    <iso25964:identifier>L8</iso25964:identifier>
 so25964:PreferredTerm>
</iso25964:ThesaurusConcept>
```

Thesaurus Array (2/2)

- Has a unique identifier
- Can be ordered or not
- The unique parent referenced by identifier, is either of:
  - A Concept
  - A (super-) Array
- The members are a combination of 1 or more
  - Concepts
  - (sub-) Arrays



The value 'true' indicates that the concepts and sub-arrays of the array are ordered. This imposed order is reflected by the XML document order of the XML document instance.

This is an optional attribute of ThesaurusConcept, ThesaurusArray and ConceptGroup (see 12.1.3 and 12.2.5.2).
If the thesaurus uses an expressive notation, then applying it to node labels will allow them to be shown in the correct place in hierarchical displays.
If there is no expressive notation, some other means should be found of outputting the node labels correctly in the display, such as a sort code attribute that is not displayed to users. In this event, the same attribute will be required at the display level for preferred terms.
The system of notation used for ConceptGroup may be quite distinct from that used for ThesaurusConcept, and one of these systems may be present without the other.

@language should be given as an alpha-2 code selected from ISO 639-1 if present in that list, or an alpha-3 code from ISO 639-2 if not.
These codes may be extended where necessary with the additional codes described in RFC 4646[45] and listed in the IANA subtag registry[35] (see 12.4.5).

The identifier of the thesaurus concept under which this array appears.
(case that this array is not a sub-array)

The identifier of the thesaurus array under which this thesaurus sub-array appears.
(case that this array is a sub-array)

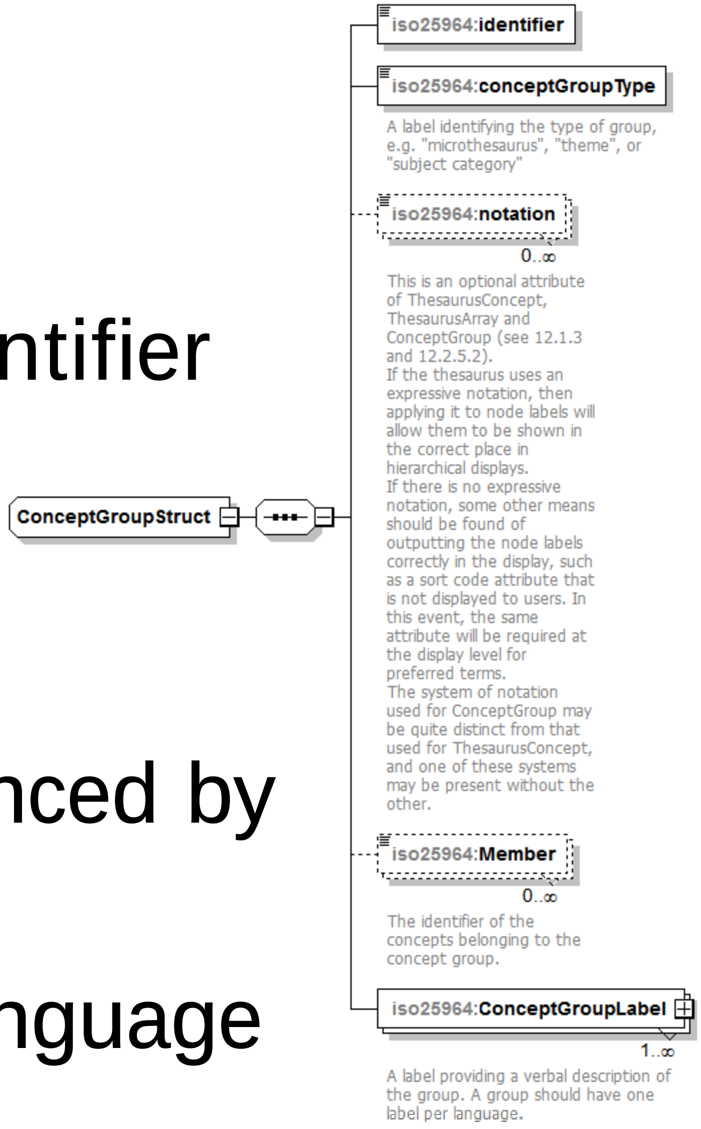
If the thesaurus array is ordered the member identifiers are listed in the order required for the array.

The identifier of a member concept of this thesaurus array.

The identifier of a sub-array of this Thesaurus Array.

Concept Group Definition (1/2)

- Each group has a unique identifier
- A group has one type
  - micro-thesaurus, theme, subject category, ...
- Member concepts are referenced by identifier
- There is a group label per language



A label identifying the type of group, e.g. "microthesaurus", "theme", or "subject category"

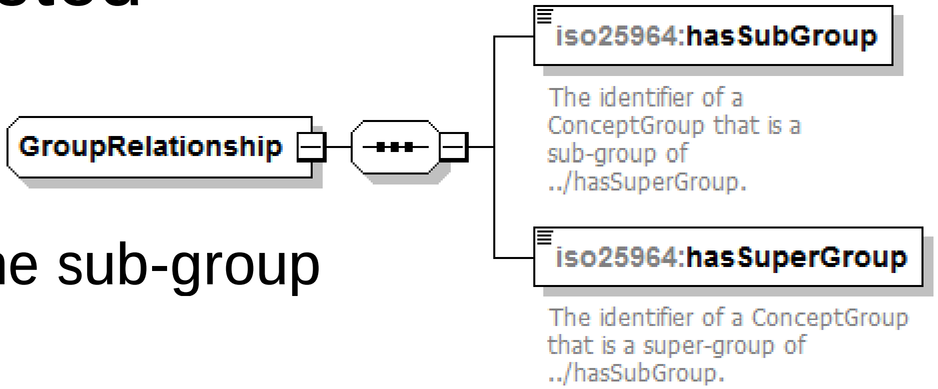
This is an optional attribute of ThesaurusConcept, ThesaurusArray and ConceptGroup (see 12.1.3 and 12.2.5.2).
If the thesaurus uses an expressive notation, then applying it to node labels will allow them to be shown in the correct place in hierarchical displays.
If there is no expressive notation, some other means should be found of outputting the node labels correctly in the display, such as a sort code attribute that is not displayed to users. In this event, the same attribute will be required at the display level for preferred terms.
The system of notation used for ConceptGroup may be quite distinct from that used for ThesaurusConcept, and one of these systems may be present without the other.

The identifier of the concepts belonging to the concept group.

A label providing a verbal description of the group. A group should have one label per language.

Concept Group sub-groups (2/2)

- Typically, groups do not form a hierarchy.
- Groups can be nested
- Relationship
  - hasSubGroup
    - the identifier of the sub-group
  - hasSuperGroup
    - the identifier of the super-group
  - a member of a sub-group is also a member of the super-group

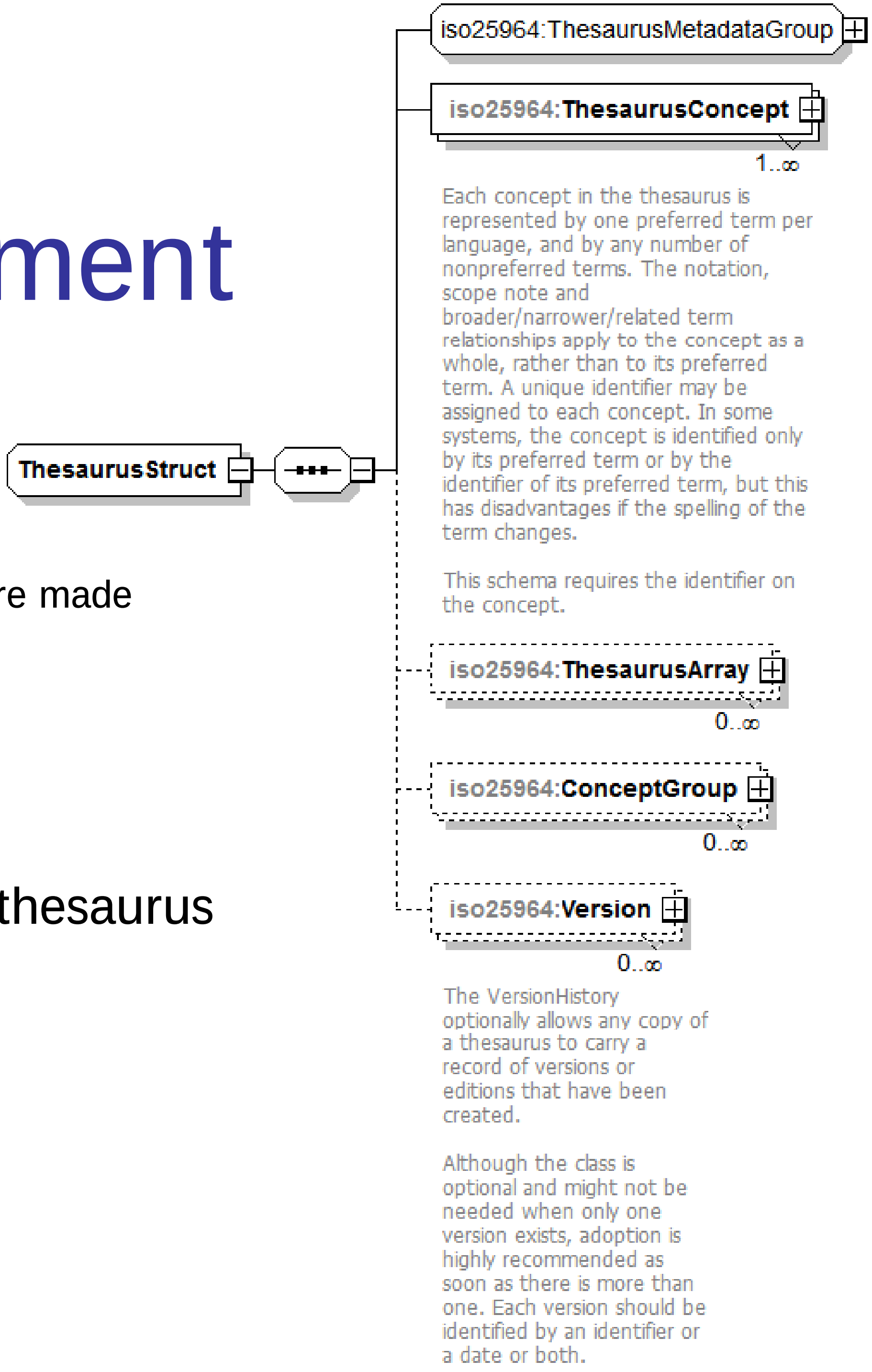


The identifier of a ConceptGroup that is a sub-group of ../hasSuperGroup.

The identifier of a ConceptGroup that is a super-group of ../hasSubGroup.

Thesaurus element

- The metadata sheet
  - identifier
  - dc:language
    - list of all languages terms are made available in
  - dc:coverage
  - dc:title
  - dc:relation
  - ....
- Basic building blocks of the thesaurus
  - ThesaurusConcept
  - ThesaurusArray
  - ConceptGroup
- Version



Each concept in the thesaurus is represented by one preferred term per language, and by any number of nonpreferred terms. The notation, scope note and broader/narrower/related term relationships apply to the concept as a whole, rather than to its preferred term. A unique identifier may be assigned to each concept. In some systems, the concept is identified only by its preferred term or by the identifier of its preferred term, but this has disadvantages if the spelling of the term changes.

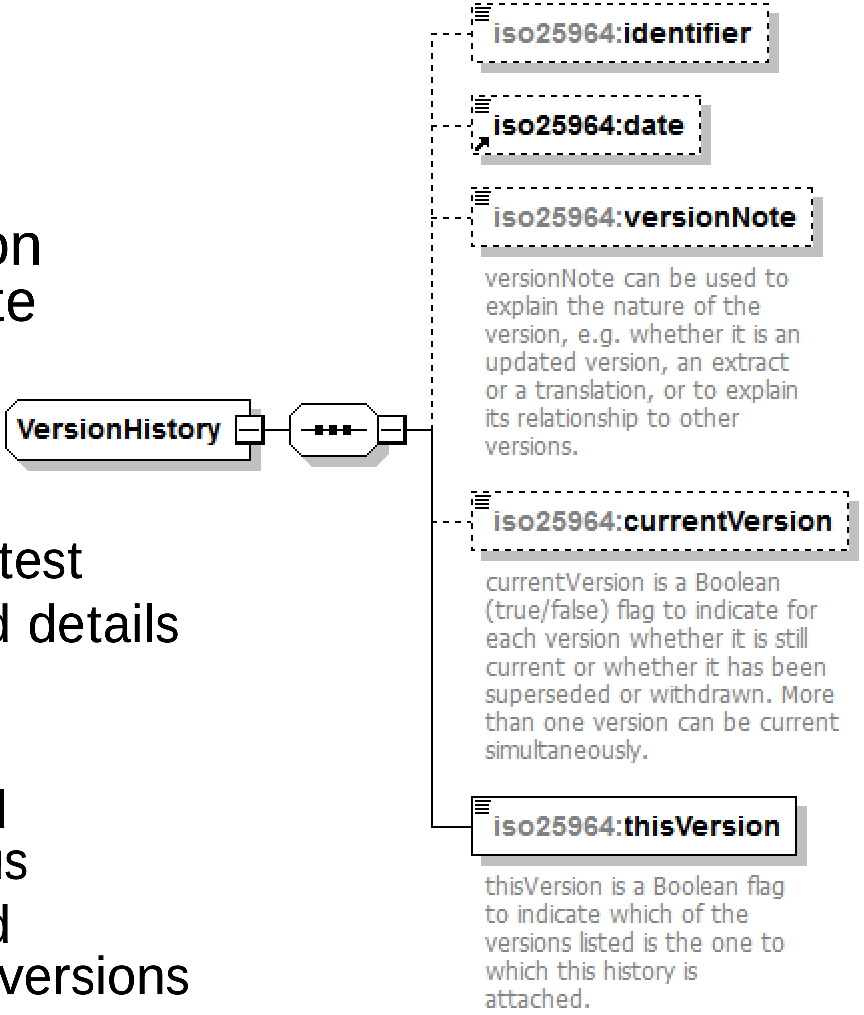
This schema requires the identifier on the concept.

The VersionHistory optionally allows any copy of a thesaurus to carry a record of versions or editions that have been created.

Although the class is optional and might not be needed when only one version exists, adoption is highly recommended as soon as there is more than one. Each version should be identified by an identifier or a date or both.

Thesaurus Version History

- Each record details a version described in the versionNote
- Each version has a date
- currentVersion
  - True: if the version record details the latest and greatest
  - False: if the version record details an older version
- thisVersion
  - True: if this version record details the linked thesaurus
  - False: if the version record pertains to other thesauri versions than the linked



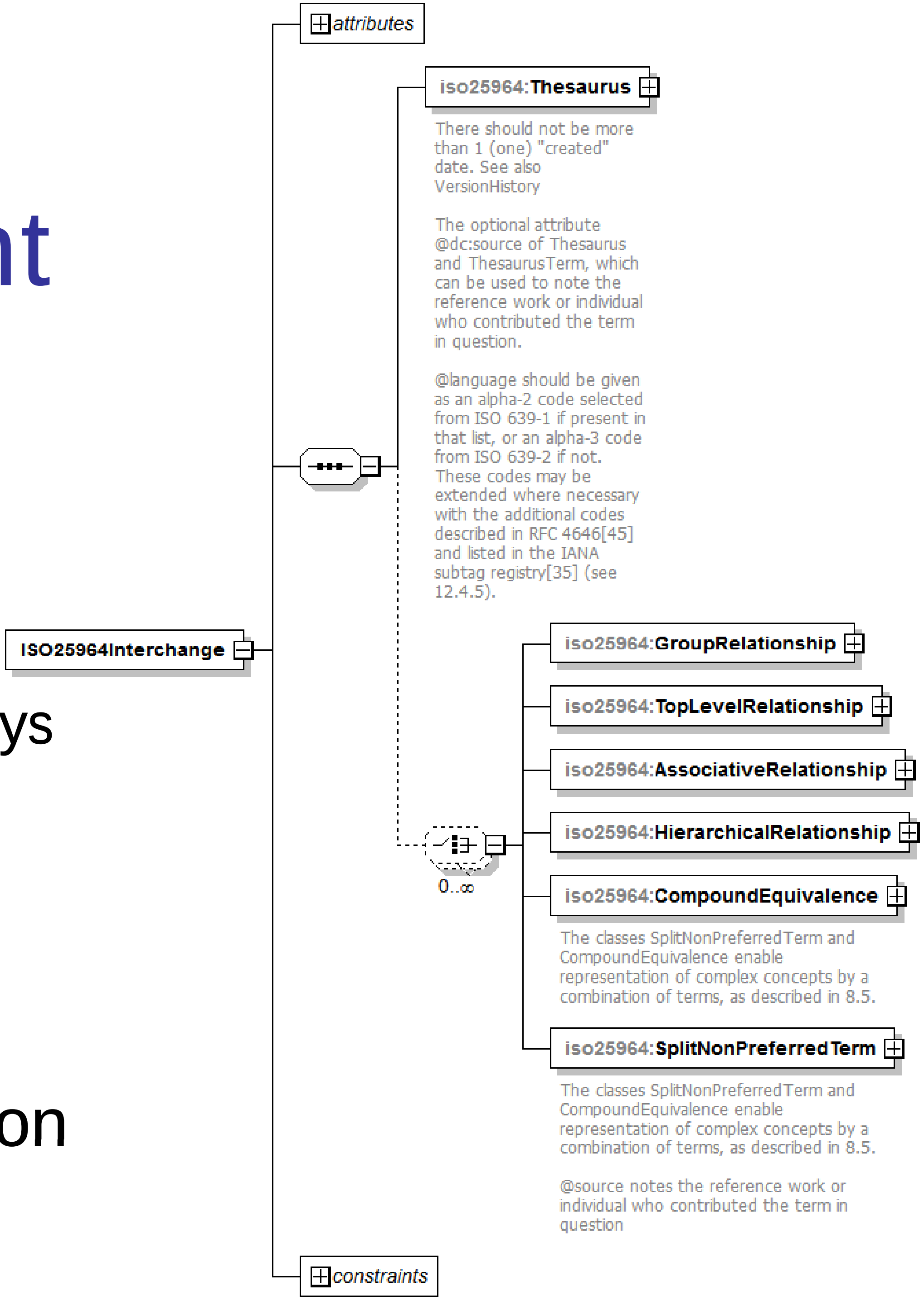
versionNote can be used to explain the nature of the version, e.g. whether it is an updated version, an extract or a translation, or to explain its relationship to other versions.

currentVersion is a Boolean (true/false) flag to indicate for each version whether it is still current or whether it has been superseded or withdrawn. More than one version can be current simultaneously.

thisVersion is a Boolean flag to indicate which of the versions listed is the one to which this history is attached.

Root element

- Thesaurus
  - details on next slide
    - metadata sheet
    - concepts, groups, arrays
- Relationships
  - of Concept Group
  - of Concepts
  - of split (compound) non Preferred Terms


attributes
iso25964:Thesaurus
ISO25964Interchange
iso25964:GroupRelationship
iso25964:TopLevelRelationship
iso25964:AssociativeRelationship
iso25964:HierarchicalRelationship
iso25964:CompoundEquivalence
iso25964:SplitNonPreferredTerm
0..∞
constraints

There should not be more than 1 (one) "created" date. See also VersionHistory

The optional attribute @dc:source of Thesaurus and ThesaurusTerm, which can be used to note the reference work or individual who contributed the term in question.

@language should be given as an alpha-2 code selected from ISO 639-1 if present in that list, or an alpha-3 code from ISO 639-2 if not. These codes may be extended where necessary with the additional codes described in RFC 4646[45] and listed in the IANA subtag registry[35] (see 12.4.5).

The classes SplitNonPreferredTerm and CompoundEquivalence enable representation of complex concepts by a combination of terms, as described in 8.5.

The classes SplitNonPreferredTerm and CompoundEquivalence enable representation of complex concepts by a combination of terms, as described in 8.5.

@source notes the reference work or individual who contributed the term in question

Envisioned extensions

- Distribution of thesaurus updates
- Complex compound relationships

XML Schema versus SKOS (1/2)

- XML and XML Schema
  - Pro
    - Reasonably well know
    - Stable and general available toolset with IDE support
      - xml, xslt, xquery
    - Strong typing
    - Integrity constraints
    - Covers all standardized features
  - Con
    - XML structure limits flexibility to order elements
    - Limited flexibility in constraints (e.g. non-cyclic graphs)
    - Limited extensibility
    - No standard internet access protocol
  - Usage
    - Exchange between partners with an established SLA

XML Schema versus SKOS (2/2)

- SKOS
  - Pro
    - Reasonably well know
    - Strong typing
    - Extensibility
    - Powerful constraint languages (SPARQL, OWL, SWRL, RIF ...)
    - Graph model provides flexible specification without ordering
    - Limited flexibility in constraints (e.g. non-cyclic graphs)
    - Standard internet access protocol (SPARQL, content negotiation)
  - Con
    - Limited toolset
      - validation, transformation, IDE
    - Needs extensions to cover all standardized features
  - Usage
    - Internet publishing, L(O)D publishing