

# GUIDELINES FOR ENCODING DOMAIN LABELS FOR LINKED DATA LEXICAL RESOURCES IN RDF

Anas Fahad Khan, Ana Salgado, Rute Costa, Margarida Ramos, Bruno Almeida, Sara Carvalho and Raquel Silva



## DOMAIN LABELS – AN INTRODUCTION

- In lexicography, the term **domain label** commonly used to denote a **usage label** assigned to a sense or a whole entry and which serves as a ‘marker which identifies the specialised field of knowledge in which a lexical unit is mainly used’ (Salgado, Costa & Tasovac, 2019).
- These labels are used ‘para señalar el léxico temáticamente especializado, en contraposición al léxico común’ [to signal the thematically specialised lexicon in contrast to the common lexicon] (Estopà, 1998, p. 1) and are generally expressed in the form of **abbreviations** representing **individual domains**
- Domain labels can be organised in **taxonomies** or **thesauri**, which can help make lexicons easier to **navigate** and to **query**. Although such labels play an essential role in lexical resources, and especially in lexicographic resources, so far there has not been much work on **modelling** them in **linked data lexicons** in a way that better exploits the possibilities of the Semantic Web stack (see however Almeida et al., 2022).
- To help remedy this state affairs, we propose a **series of guidelines for encoding domain label information in RDF** principally using three linked data vocabularies, namely **OntoLex-Lemon**, **SKOS**, and **lexicog**.

## ONTOLEX-LEMON AND DOMAIN LABELS

- While OntoLex-Lemon did not retain these properties, the OntoLex-Lemon guidelines instead suggest the use of the **dct:subject** property to specify: *under which conditions (context, register, domain, etc) it is valid to regard the lexical entry as having the ontological entity as meaning.*
- The same guidelines also recommend the use of the **ontolex:usage** property, defined as specifying the *usage conditions or pragmatic implications when using the lexical entry to refer to the given ontological meaning*
- This property has the domain **ontolex:LexicalSense** and the range **rdfs:Resource**. Moreover, the lexicog vocabulary defines a series of subproperties of **ontolex:usage** including **lexinfo:domain** which is defined as: *usage marker which identifies the specialized field of knowledge in which a lexical unit is mainly used.*
- Ontolex therefore offers us a way of marking a lexical entry as belonging to a certain domain and a way of specifying that a specific sense of an entry is associated with a particular domain.
- When it comes to encoding the domain label itself, we suggest encoding it as an instance of the SKOS class **Concept** and using the **skos:narrower** and **skos:broader** relations to encode the relations between different domains.

## BEST PRACTISES FOR ENCODING DOMAIN LABELS

1. Domain labels should be encoded as individuals of the class **skos:Concept**. Hierarchical relationships between individual domain labels should be encoded using the **skos:narrower** and **skos:broader** properties. In the case of retrodigitised and non-native-born dictionaries, it may be that the same domain label is not consistently encoded using the same string; in such situations, we recommend using **skos:prefLabel** and **skos:altLabel** to list the different versions of the same label (with the former being used to encode the version(s) found in the front matter and the latter its variants).
2. In case the whole entry is marked as (or interpreted by the encoder as) belonging to a given domain we recommend encoding this information using **dcterms:subject** with the entry as subject and the relevant domain label (encoded as **skos:Concept**, see above) as object.
3. In case a single sense is marked as (or is interpreted by the encoder as) belonging to a domain, we recommend using **lexinfo:domain** with the entry as subject and the relevant **skos:Concept** as object.
4. In other cases where any other part of the entry is marked with a domain label, once again we recommend the use of **dcterms:subject**.

## EXAMPLE

**cristalografia** [kɾiʃtɐlɔgɾɨˈfiɐ]

**Entrada validada**

*nome feminino*

**MINERALOGIA** ciência que estuda os cristais, considerando aspetos tais como o seu crescimento, a estrutura interna e as propriedades físicas decorrentes da regularidade dessa estrutura, em particular, as formas que apresentam, cuja simetria utiliza como método de classificação e de descrição

**ETIMOLOGIA** Do grego κρύσταλλος, ‘cristal’ + sufixo **-grafia**

This entry is for the Portuguese lexical unit *cristalografia* 'crystallography' and comes from the Academia dictionary. As the following figure shows, this entry has one sense which is marked with the label **MINERALOGIA** referring to the domain of mineralogy. Some additional information relevant to this example is that the domain of **MINERALOGIA** is a subdomain of the **GEOLOGIA** 'geology' domain in the Academia dictionary subject hierarchy that belongs to **CIÊNCIAS DA TERRA** 'earth sciences' superdomain. We can represent these domains and their interrelations as follows using the SKOS vocabulary:

```
<http://example.org/class/mineralogia> rdf:type skos:Concept;
  skos:prefLabel "mineralogia"@pt;
  skos:prefLabel "minerology"@en;
  skos:narrower <http://example.org/class/geologia> .
<http://example.org/class/geologia> rdf:type skos:Concept;
  skos:prefLabel "geologia"@pt;
  skos:prefLabel "geology"@en;
  skos:narrower <http://example.org/class/ciencias_da_terra> ;
  skos:broader <http://example.org/class/mineralogia> .
<http://example.org/class/ciencias_da_terra> rdf:type
skos:Concept;
  skos:prefLabel "ciencias da terra"@pt;
  skos:prefLabel "earth sciences"@en;
  skos:broader <http://example.org/class/mineralogia> .
```

In the entry itself, we link the (single) sense of the entry for *cristalografia* to the domain `<http://example.org/class/mineralogia>` via the **lexinfo:domain** property.

```
http://example.org/class/DLP_cristalografia> a ontolex:LexicalEntry
;
  lexinfo:etymology
    [ rdf:value "Do grego κρύσταλλος cristal + sufixo
      -grafia"@pt ] ;
  lexinfo:gender lexinfo:feminine ;
  lexinfo:partOfSpeech lexinfo:noun ;
  ontolex:canonicalForm [
    ontolex:phoneticRep "kɾiʃtɐlɔgɾɨ'fiɐ"@pt ;
    ontolex:writtenRep "cristalografia"@pt
  ] ;
  ontolex:sense [
    lexinfo:domain <http://example.org/class/mineralogia>;
    skos:definition ""ciência que estuda os cristais,
considerando aspetos tais como o seu crescimento, a
estrutura interna e as propriedades físicas
decorrentes da regularidade dessa estrutura,
em particular, as formas que apresentam, cuja
simetria utiliza como método de
classificação e de descrição""@pt
  ] .
```











