

# Vocabulary language metadata

The metadata associated with every vocabulary in RVA includes a list of languages in which the vocabulary's data is available. When you publish a vocabulary in RVA, you must specify at least one such language. This page explains how that's done, and how the language metadata is used by RVA.

## Languages, language tags, and descriptions

The main thing to know is that the languages of your vocabulary are almost certainly supported as valid metadata. As mentioned below, when entering metadata in RVA, the most commonly-used languages are selectable from a simple dropdown, but languages not in that list can also be entered.

The available selection of languages is based on [IETF language tags](#), which are determined by the IETF document [BCP 47](#). This selection is consistent with the language tags used in `RDF langStrings`; therefore, in particular, all valid RDF language tags are supported. There is one main restriction of RVA's support: private use subtags (e.g., `qaa`) are not permitted.

BCP 47 incorporates and subsumes the lists of languages included in ISO 639 and other standards. In many of the most common cases, ISO 639-1 language tags are also valid BCP 47 tags. Here are some examples of valid BCP 47 languages: their descriptions and their tags:

Description	BCP 47 tag
English	<code>en</code>
French	<code>fr</code>
French (Canada)	<code>fr-CA</code>
Chinese	<code>zh</code>
Chinese (Han (Simplified variant))	<code>zh-Hans</code>
Ngaanyatjarra	<code>ntj</code>

For detailed advice about selecting BCP 47 language tags, please see the W3C pages [Language tags in HTML and XML](#) and [Choosing a Language Tag](#). You may find Richard Ishida's online, interactive [BCP47 language subtag lookup](#) tools helpful. If you would like further assistance in selecting language metadata for your vocabulary, please contact [services@ardc.edu.au](mailto:services@ardc.edu.au).

## Entering language metadata

RVA's metadata entry page, also known as the CMS, has a section for entering language metadata. You may enter as many languages as you wish.

The first language that you specify is treated specially. It is referred to as the vocabulary's "primary language". The subsequent languages in the list are known as the vocabulary's "other languages". See the section "How vocabulary language metadata is used" below for information about the significance of the primary language.

If you use RVA's ability to import a vocabulary from PoolParty, the pre-filling of the metadata entry page will automatically assign the PoolParty project's own primary language as the primary language. (Note that this pre-filling only happens once, at the moment when you select the PoolParty project. If you subsequently change the language metadata of your PoolParty project, you will need to make the corresponding changes manually in RVA's metadata entry page.)

The most commonly-used languages are available from a dropdown, but you can also select languages not already shown. To do this, click the dropdown to open the text input, then type in the language's BCP 47 tag and press Enter. The tag will be looked up, and if it is a valid tag, the language tag will be canonicalised, and the language's description will be displayed. (Canonicalisation refers to the BCP 47 process by which certain tag values are simplified. For example, the tag `en-Latn` is canonicalised to just `en`.)

## The fine details of language tag validation

This section is provided for completeness only; you almost certainly don't need to worry about these very fine details.

The validation of language tags conforms to the rules spelled out in BCP 47, with the following caveats and exceptions:

- A tag will be rejected as invalid, if it contains:
  - a deprecated subtag for which the IANA Language Subtag Registry does not specify a preferred value,

- a subtag marked as "private use",
  - *only* extensions and private use subtags, even if it would otherwise be well-formed
- If a tag contains a subtag for which there is a preferred value, the subtag will be canonicalised as that preferred value.
- If a tag is grandfathered, and there is a preferred value, the tag is canonicalised as that preferred value; otherwise, it is rejected as invalid (because it is deprecated).
- If a tag is redundant, and there is a preferred value, it is canonicalised as the preferred value; otherwise, it is canonicalised using the individual subtags.
- Because satisfying RFC 5646 section 4.1, point 6, on the validity and ordering of variants, is particularly complicated, some of those validity checks are not yet performed. This means that some combinations of variants are accepted instead of being rejected.

## How vocabulary language metadata is used

### Search index

Each vocabulary's languages are indexed, and are available in search results as facet values. Please note: at present, the search index contains only each language's "language subtag", which means the first, main component of the language description. For example, the language tags `en` and `en-AU` are both indexed only as "English".

### Browse visualisation

If the current version of a vocabulary contains SKOS resources, so that there is a browse visualisation displayed on the vocabulary's view page, the primary language is taken into account when determining which label is shown on the nodes of the display. But note: an instance of a resource's label that is not language-tagged takes priority over all that are.

Example: if the vocabulary's primary language is German (`de`), and the vocabulary data contains these labels for a concept:

```
my:one skos:prefLabel "Eins"@de , "One"@en .
```

then the label chosen for the display will be `Eins`.

But if the data is:

```
my:one skos:prefLabel "Eins"@de, "One"@en, "One" .
```

then the label chosen for the display will be `One`, because a label *without* a language tag takes priority over all language-tagged labels for the same resource, even one for the primary language.