

General Guidelines:

- Content **MUST** be at least provided in English, exceptions are only acceptable in case of names (persons, organisations etc.). The content **MAY** be provided in other languages too.
- Attribute *xml:lang* **MUST** be used for all non-English content
- Existing and widely-used vocabularies **SHOULD** be used
- for used vocabularies the exact spelling specified in the vocabulary **MUST** be used
- For names (persons, organisations etc.) standardised variants **SHOULD** always be used
- Atomic values **MUST** be used (instead of a single 'deu, eng' → 'deu', 'eng')
- Different values **SHOULD** be orthogonal, they **SHOULD NOT** have semantic overlap
- Unspecified/Underspecified values **SHOULD** be avoided (e.g. instead of 'multilingual' give all valid values)
- Values supposed to be used for faceted search **SHOULD** be relatively short (i.e. single words, short phrases)
- For some facets (like genre, subject, keywords) different vocabularies **MAY** be used. These **SHOULD** be consistently used inside the same collection.

Useful vocabularies:

Facet	Vocabulary	Useful CMDI Component
Collection	-	-
Language	ISO 639-3 (language) or ISO 639-5 (language families) code or corresponding English name	cmdi-language <sup>1</sup> ( <a href="http://clarin.eu:cr1:c_1271859438111">clarin.eu:cr1:c_1271859438111</a> ), iso-language-639-3 <sup>1</sup> ( <a href="http://clarin.eu:cr1:c_1271859438110">clarin.eu:cr1:c_1271859438110</a> ), iso-languagefamily-639-5 <sup>13</sup> ( <a href="http://clarin.eu:cr1:c_1271859438108">clarin.eu:cr1:c_1271859438108</a> )
ResourceClass	DC-3806	-
Modality	NaLiDa Modalities <sup>2</sup>	ModalityInfo ( <a href="http://clarin.eu:cr1:c_1290431694513">clarin.eu:cr1:c_1290431694513</a> )
Continent	Geonames convention ( <a href="http://Geonames Webservices">Geonames Webservices</a> ) or corresponding English name (DC-2531 <sup>12</sup> )	iso-continent <sup>1</sup> ( <a href="http://clarin.eu:cr1:c_1271859438105">clarin.eu:cr1:c_1271859438105</a> )
Genre	-	-
Country	ISO 3166-1 code or their corresponding English names <b>Jörg Knappen</b>	iso-country <sup>1</sup> ( <a href="http://clarin.eu:cr1:c_1271859438104">clarin.eu:cr1:c_1271859438104</a> )
Subject	-	-
Organisation	-	-
Format	all MIME-Types	cmdi-mimetype ( <a href="http://clarin.eu:cr1:c_1271859438106">clarin.eu:cr1:c_1271859438106</a> )
Keywords	-	-
Year (bzw.	W3C DateTime	-

Facet	Vocabulary	Useful CMDI Component
beliebige Zeitangaben)		
LiveCycleStatus	DC-3818	-
Time Coverage	Open Date Range Format	-

<sup>1</sup> Definition of used ISOcat category problematic

<sup>2</sup> Definition only in CMDI component (=not a complex/closed datcat in ISOcat)

<sup>3</sup> Incomplete definition in component (missing conceptlinks)