

Gruppe Vokabular

VLO Taskforce

- Inhalte immer *auch* in Englisch (Ausnahmen möglich), Nutzung von xml:lang zumindest für nicht-Englische Texte
- Möglichst vorhandene und verbreitete Vokabulare nutzen; Nutzung 1:1 Schreibweise
- Möglichst standardisierte Verwendung von Namen (Personen, Organisationen etc.)

- Immer Einzelwerte je Element („deu, eng“ → „deu“, „eng“)
- Verschiedene Elemente sollten möglichst von ihrer Bedeutung orthogonal sein, keine inhaltliche Überschneidung
- Möglichst keine unspezifischen Elemente, sondern explizite Angaben („multilingual“ → „deu“, „eng“, „fra“)

- Werte die für facettierte Suche gedacht/geeignet sind, möglichst kurz (Einzelwörter, kurze Wortgruppen)

- Hauptproblem: weiterhin keine Definition der Facetten („country“, „language“, „genre“)
- Derzeit nur
 - Explizite (ad hoc) Definition durch Mengen von Conceptlinks
 - vage (vermutlich häufig falsche) Definition in Volltexten

Facette	Vokabulare	Empfohlene CMDI-Komponente
Collection	-	-
Language	Language Codes nach ISO 639-3 (Sprachen) bzw. ISO 639-5 (Sprachfamilien) bzw. entsprechender englischer Name	cmdi-language ¹ (clarin.eu:cr1:c_1271859438111), iso-language-639-3 ¹ (clarin.eu:cr1:c_1271859438110), iso-languagefamily-639-5 ¹³ (clarin.eu:cr1:c_1271859438108)
ResourceClass	Stark abhängig von der Facetten-Definition: Dublin Core Type (DCMI Type Vocabulary), DC-3806, OLAC Linguistic Data Type	-
Modality	NaLiDa Modalities ² (DC-2490 ²)	ModalityInfo (clarin.eu:cr1:c_1290431694513)
Continent	Codes entsprechend Geonames Konvention (Geonames Webservices) bzw. entsprechender englischer Name (DC-2531 ¹²)	iso-continent ¹ (clarin.eu:cr1:c_1271859438105)

Facette	Vokabulare	Empfohlene CMDI-Komponente
Genre	-	-
Country	Codes nach ISO 3166-1 bzw. englischer Name entsprechend Standard	iso-country ¹ (clarin.eu:cr1:c_1271859438104)
Subject	-	-
Organisation	-	-
Format	Beliebiger MIME-Type	cmdi-mimetype (clarin.eu:cr1:c_1271859438106)
Keywords	-	-
Year (bzw. beliebige Zeitangaben)	W3C DateTime	-

- Mögliche zukünftige Facetten:

Facette	Vokabulare	Empfohlene CMDI-Komponente
LiveCycleStatus	DC-3818 („life cycle status“)	-
Time Coverage	Open Date Range Format (→ Jörg)	-

- ¹ Definition verwendeter ISOcat-Kategorie(n) problematisch
- ² Definition nur über CMDI Komponente (=offene DatCat in ISOcat)
- ³ Unvollständige Definition der Komponente (fehlende Conceptlinks)