

Daniel Broeder, Freddy Offenga - IMDI Metadata Set 3.0

Based on the experiences and on a broad discussion process including field linguists, corpus linguists and language engineers, the IMDI set 3.0 [1] was designed as part of the INTERA and DOBES projects and is available as an XML-Schema. It was adapted to simplify the content description and the artificial distinction between collectors and other participants - probably influenced by Dublin Core - was removed.

Three major extensions were applied: First, it is now possible to describe written resources that are not annotations or descriptions. This was necessary, since most language collections contain written resources in the form of field notes, sketch grammars, phoneme descriptions and others more. Second, as a consequence of long discussions with participants of the MILE lexicon initiative, it is now possible to describe lexicons with a specialized set of descriptor elements. Third, it is now possible to define and add project-specific profiles. In the earlier version IMDI supported already the possibility of extensions at various levels in the form of user defined category-value pairs, i.e., the user was able to define a private category and associate values with it. This feature was used by individuals and also projects to include special descriptors, however, these descriptors were not fully supported by the IMDI tools. In the new version projects or sub-domains such as the Dutch Spoken Corpus respectively the Sign Language community can define a set of important categories and these are supported while editing or searching.

Therefore, IMDI exists of its core definitions that have to be stable to assure users that their work will be exploitable even after many years and of sub-community specific extensions, which nevertheless are result of discussion processes.

[1] Detailed description of the IMDI 3.0 metadata elements:

http://www.mpi.nl/IMDI/documents/Proposals/IMDI_Metadata_3_0_4.pdf

[2] IMDI Web-site: <http://www.mpi.nl/IMDI/>