
**Language resource management —
Component metadata infrastructure
(CMDI) —**

**Part 2:
Component metadata specification
language**

*Gestion des ressources linguistiques — Composante infrastructure de
métadonnées (CMDI) —*

Partie 2: Composante linguistique spécifique aux métadonnées



This document is a preview generated by EMS



COPYRIGHT PROTECTED DOCUMENT

© ISO 2019

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	iv
Introduction	v
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
3.1 General terms	1
3.2 CMDI	3
3.3 XML	5
4 Notational and XML namespace conventions	7
5 Structure of CMDI instances	8
5.1 General structure	8
5.2 The main structure	9
5.3 The <Header> element	10
5.4 The <Resources> element	11
5.4.1 General structure of the <Resources> element	11
5.4.2 The list of resource proxies	11
5.4.3 The list of journal files	12
5.4.4 The list of relations between resource files	13
5.5 The <IsPartOfList> element	15
5.6 The CMD components	15
6 CCSL (CMDI Component Specification Language)	17
6.1 General structure of the CCSL	17
6.2 CCSL header	19
6.3 CMD specification	20
6.4 Definition of CMD elements	21
6.5 CMD attribute definition	23
6.6 Value schemes for CMD elements and CMD attributes	24
6.7 Cue attributes	26
7 CMD	27
7.1 Transformation of CCSL into a CMD profile schema definition	27
7.2 General properties of the CMD profile schema definition	27
7.3 Interpretation of CMD specifications in the CCSL	27
7.3.1 General structure of CMD specifications	27
7.3.2 Document structure prescribed by the CMD profile schema	28
7.4 Interpretation of CMD element definitions in the CCSL	28
7.5 Interpretation of CMD attribute definitions in the CCSL	29
7.6 Content model for CMD elements and CMD attributes in the schema definition	30
Bibliography	31

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

A list of all parts in the ISO 24622 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Many researchers, from the humanities and other domains, have a strong need to study resources in close detail. Nowadays more and more of these resources are available online. To be able to find these resources, they are described with metadata. These component metadata (CMD) instances are collected and made available via central catalogues. Often, resource providers want to include specific properties of a resource in their metadata to provide all relevant descriptions for a specific type of resource. The purpose of catalogues tends to be more generic and addresses a broader target audience. It is hard to strike the balance between these two ends of the spectrum with one metadata schema, and mismatches can negatively impact the quality of metadata provided. The goal of the component metadata infrastructure (CMDI) is to provide a flexible mechanism to build resource specific metadata schemas out of shared components and semantics^{[14][15]}.

In CMDI the metadata lifecycle starts with the need of a metadata modeller to create a dedicated metadata profile for a specific type of resource. Modellers can browse and search a registry for components and profiles that are suitable or come close to meeting their requirements. A component groups together metadata elements that belong together and can potentially be reused in a different context. Components can also group other components. Existing component registries, e.g., the CLARIN (common language resources and technology infrastructure) Component Registry^[16], might already contain any number of components. These can be reused as they are, or be adapted by modifying, adding or removing some metadata elements and/or components. Also completely new components can be created to model the unique aspects of the resources under consideration. All the needed components are combined into one profile specific for the type of resources. Any component, element and value in such a profile may be linked to a semantic description — a *concept* — to make their meaning explicit^[21]. These semantic descriptions can be stored in a semantic registry, e.g., the CLARIN Concept Registry^[17]. In the end metadata creators can create records for specific resources that comply with the profile relevant for the resource type, and these records can be provided to local and global catalogues^[22].

CMDI has originally been developed in the context of the European CLARIN infrastructure initiative with input from other initiatives and experts. Already in its preparatory phase, which started in 2007, the infrastructure needed flexibility in the metadata domain as it was confronted with many types of resources that had to be accurately described. For Version 1.0 a toolkit^[20] was created, consisting of the XML schemas and XSLT stylesheets to validate and transform components, profiles and records. Version 1.1 included some small changes and has seen small incremental backward compatible advances since 2011. This version has been in use, new developments and the development of this document resulted in Version 1.2^[18]. Also CMDI has seen a growing number of tools and infrastructure systems that deal with its records and components and rely on its shared syntax and semantics.

In ISO 24622-1, the component metadata model has been standardized. This document is compliant with ISO 24622-1, and also extends and constrains it at various places (see also the red parts in the UML class diagram in [Figure 1](#)):

- support for attributes on both components and elements is added,
- a profile is limited to one root component, and
- an element always belongs to a specific component.

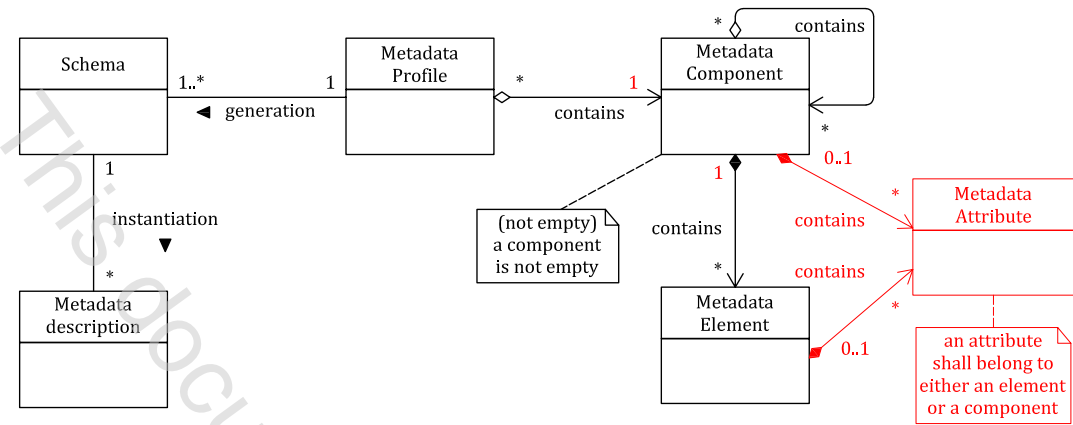


Figure 1 — Component metadata model and its extensions

Language resource management — Component metadata infrastructure (CMDI) —

Part 2: Component metadata specification language

IMPORTANT — The electronic file of this document contains colours which are considered to be useful for the correct understanding of the document. Users should therefore consider printing this document using a colour printer.

1 Scope

The component metadata lifecycle needs a comprehensive infrastructure with systems that cooperate well together. To enable this level of cooperation this document provides in depth descriptions and definitions of what CMDI records, components and their representations in XML look like.

This document describes these XML representations, which enable the flexible construction of interoperable metadata schemas suitable for, but not limited to, describing language resources. The metadata schemas based on these representations can be used to describe resources at different levels of granularity (e.g. descriptions on the collection level or on the level of individual resources).

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <http://www.iso.org/obp>
- IEC Electropedia: available at <http://www.electropedia.org/>

3.1 General terms

3.1.1 concept

unit of knowledge created by a unique combination of characteristics

[SOURCE: ISO 1087:—¹⁾, 3.2.3, modified — Note 1 to entry and Note 2 to entry have been deleted.]

3.1.2 concept link

reference from a *CMD profile* (3.2.11), *CMD component* (3.2.3), *CMD element* (3.2.5), *CMD attribute* (3.2.2) or a value in a *controlled vocabulary* (3.1.4) to an entry in a *semantic registry* (3.1.11) via a *Uniform Resource Identifier* (3.1.13)

Note 1 to entry: Typically a concept link is provided as a *persistent identifier* (3.1.9).

1) Revision of ISO 1087:2000 under preparation. Stage at the time of publication: ISO/FDIS 1087:2019.