
**Management of terminology
resources — Data categories —**

**Part 1:
Specifications**

*Gestion des ressources terminologiques — Catégories de données —
Partie 1: Spécifications*





COPYRIGHT PROTECTED DOCUMENT

© ISO 2022

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
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
4 Data categories and data category specifications	4
5 General recommendations for data category specifications	4
6 Detailed requirements for documenting a data category in a DCR	5
6.1 Identifiers and names.....	5
6.1.1 A unique and stable mnemonic identifier.....	5
6.1.2 A persistent identifier (PID).....	5
6.1.3 A unique canonical data category name.....	5
6.1.4 Language-specific data category names.....	5
6.2 Conceptual domains, data category selections and data category types.....	6
6.3 Data elementarity.....	6
6.4 Profiles.....	6
7 Referencing data categories	7
8 Harmonizing and vetting data categories	7
9 Management	8
Annex A (informative) Structure of a data category specification	9
Bibliography	12

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 3, *Management of terminology resources*.

This first edition of ISO 12620-1, together with ISO 12620-2:2022, cancels and replaces ISO 12620:2019, which has been divided into parts and technically revised. The main changes are as follows:

- ISO 12620:2019 described procedures for defining data categories used in language resources and described requirements for maintaining a pragmatic, consensus-based repository of harmonized data category specifications for use in language resources. This document has been narrowed to focus on the structure and rationale associated with data category specifications per se.
- The sections of ISO 12620:2019 that dealt with the creation and maintenance of data category repositories have been moved to ISO 12620-2.

A list of all parts in the ISO 12620 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

Data associated with language resources are identified, collected, managed and stored in a wide variety of environments. Data appearing in language resources are generalized into classes that are referred to as “data categories”. Differences in approach for developing different kinds of language resources as well as differences in technical environments inevitably lead to variations in data category definitions and data category names. The use of uniform data category names and definitions employed in resources within the same linguistic domain (e.g. among terminology resources, lexical resources, annotated text corpora) contributes to system coherence and enhances the re-usability of data. Such uniform use requires access to formal data category specifications. Defining a clear framework for specifying, managing and using data categories will increase interoperability of language resources.

The intended audience of this document is researchers and practitioners in fields of language resource management who use data categories and data category specifications.

Management of terminology resources — Data categories —

Part 1: Specifications

1 Scope

This document provides requirements and recommendations governing data category specifications for language resources. It specifies mechanisms for creating, documenting, harmonizing and maintaining data category specifications in a data category repository (DCR). It also describes the structure and content of data category specifications.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 12620-2, *Management of terminology resources — Data categories — Part 2: Repositories*

ISO 24619, *Language resource management — Persistent identification and sustainable access (PISA)*

3 Terms and definitions

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

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

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

3.1

conceptual domain

permissible content of a *data category* (3.2)

EXAMPLE In a terminology database, the data category */part of speech/* can have a conceptual domain consisting of the values */noun/*, */verb/*, */adjective/*, */adverb/*.

Note 1 to entry: The permissible content can be closed, as in the example, or subject to formal restrictions such as dates, or free text such as the conceptual domain of */definition/*. Although the latter type is not formally restricted, it is nevertheless subject to adherence to the requirements of its data category specification, i.e. it contains a true definition and not a note, example, or some other piece of information.

3.1.1

open conceptual domain

conceptual domain (3.1) that has no formal restrictions

Note 1 to entry: An open conceptual domain is frequently associated with data categories that take free text as their content, such as */definition/* or */context/*.

Note 2 to entry: Some requirements are not always machine-processable, for instance, to require that */definition/* only contain definitional information, or that a */context/* meet certain specified requirements.