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

**Part 2:
Repositories**

*Gestion des ressources terminologiques — Catégories de données —
Partie 2: Répertoires*



This document is a preview generated by ELS



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 category repositories..... | 2 |
| 4.1 Data category environments..... | 2 |
| 4.2 Requirements for a DCR..... | 3 |
| 4.3 Data category profiles..... | 3 |
| 4.4 Data category selections..... | 4 |
| 5 Roles with governance responsibilities..... | 4 |
| 5.1 Communities of practice..... | 4 |
| 5.1.1 Description..... | 4 |
| 5.1.2 Procedures for establishing a DCR..... | 4 |
| 5.2 Profile experts..... | 5 |
| 5.2.1 Description..... | 5 |
| 5.2.2 Roles and responsibilities..... | 5 |
| 5.3 Profile managers..... | 5 |
| 5.3.1 Description..... | 5 |
| 5.3.2 Roles and responsibilities..... | 5 |
| 5.4 DCR management board..... | 6 |
| 5.4.1 Description..... | 6 |
| 5.4.2 Roles and responsibilities..... | 6 |
| 6 Workflow for data category specification and maintenance..... | 6 |
| 7 Public information and interactions..... | 7 |
| Bibliography..... | 8 |

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 12620-2, together with ISO 12620-1:2022, 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. ISO 12620-1 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 this document.

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

ISO 12620-1 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) with the goal of increasing the interoperability of language resources such as terminological resources, lexicographical resources and annotated text corpora. This document specifies procedures and practices for the creation, management and maintenance of DCRs.

Interoperability of language resources is a key factor for supporting innovation and progress in various focus areas of the language industry, such as terminology management, natural language processing and annotation schemes. These areas support important sectors of the economy and social development such as global communication and trade, knowledge extraction and content management.

Researchers and software developers working with language resources benefit greatly from being able to access a trusted source of information about data categories. Providing a precise description of the data categories that are used within a given data collection allows for a quick diagnosis of its compatibility with other data collections or its suitability for use in computer processes. A DCR containing vetted data category specifications provides users with the information they need in order to implement data categories in a manner that is consistent with other users. Consequently, the interoperability of language resources is greatly enhanced.

Data category specifications are normally stored in electronic format in a specially designed database. This database is called a “data category repository (DCR)”. Today, it is essential for DCRs to be sharable for all stakeholders. See, for instance, Reference [3], a DCR for language resource descriptions named DatCatInfo.

Management of terminology resources — Data categories —

Part 2: Repositories

1 Scope

This document establishes criteria for the management of data categories for use in the creation and maintenance of language resources within a given community of practice (CoP). It defines the roles and responsibilities associated with the creation and maintenance of such repositories. It also specifies procedures to establish a governance structure for the management of a data category repository (DCR), including the addition of new data category specifications and continuous quality assurance.

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 12620-1, *Management of terminology resources — Data categories — Part 1: Specifications*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 12620-1 and the following 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

data category specification

DC specification

complete descriptive record of a data category

[SOURCE: ISO 12620-1:2022, 3.5]

3.2

data category repository

DCR

digital collection of *data category specifications* (3.1)

EXAMPLE DatCatInfo, a DCR for language resources (see Reference [3]).

Note 1 to entry: Data category repositories are used as references when specifying language resources.

[SOURCE: ISO 12620-1:2022, 3.6]