
**Assessment and benchmarking of
terminological resources — General
concepts, principles and requirements**

*Critères d'évaluation comparative des ressources terminologiques —
Concepts, principes et exigences d'ordre général*

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 23185:2009

<https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009>



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 23185:2009

<https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2009

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

| | |
|--|----|
| Foreword | iv |
| Introduction..... | v |
| 1 Scope | 1 |
| 2 Terms and definitions | 1 |
| 3 Uses of terminological resources..... | 2 |
| 4 Terminological resources..... | 3 |
| 4.1 General | 3 |
| 4.2 Model for assessment of terminological resources | 4 |
| 4.3 General usability attributes of terminological resources..... | 4 |
| 5 Guidelines for an assessment and benchmarking project | 11 |
| 5.1 Overview..... | 11 |
| 5.2 Main steps of the workflow..... | 12 |
| 5.3 Test-case assessment | 12 |
| Annex A (normative) Usability attributes of terminological resources and basic rules for measurement | 13 |
| Annex B (informative) Typical application of usability attributes as benchmarks | 17 |
| Bibliography..... | 21 |

ISO 23185:2009

<https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009>

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 23185 was prepared by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 2, *Terminographical and lexicographical working methods*.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

[ISO 23185:2009](https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009)

<https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009>

Introduction

Global society is undergoing an accelerated development towards becoming a science- and technology-driven multilingual information and knowledge society characterized by the all-pervading influence of information and communication technology (ICT). Reliable language resources (such as text and speech corpora, terminologies, computational lexicons, etc.) are essential to support the emerging knowledge and content industries. Terminology information is thus becoming a key element in all regulatory activities, as seen, for example, in technical standardization, quality management and regulation of intellectual property rights.

In the emerging semantic web, dedicated and non-dedicated browsers or web services search web-based databases and portals containing structured content (i.e. collections of content items at the level of lexical semantics). The user increasingly does not want to be overburdened with non-evaluated information, but to receive the most pertinent and reliable information for his/her purpose without missing important information. The results of automatic or semi-automatic searches, therefore, will increasingly have to be compounded and condensed by semantic analyses in order to meet user requirements.

If seemingly relevant information is found in a multitude of collections of structured content, systematic syntactic and semantic filtering, selection and evaluation processes take place. At some stage of these processes, browsers or web services have to globally distinguish between more or less pertinent and reliable terminological data as it is being collected for the sake of prioritization and optimization.

Terminological data can have many functions, the most prominent of which are

- knowledge representation (concept),
- knowledge ordering (concept classification),
- access to other kinds of structured or unstructured content, and
- means or elements of communication and knowledge transfer.

For uses and reuses such as translation, localization and content management, a systematic approach to automatic or semi-automatic assessment and benchmarking of resources or containing terminological data becomes necessary.

iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 23185:2009

<https://standards.iteh.ai/catalog/standards/sist/637c1e0c-3c34-41a7-b8fd-b7bb5129f745/iso-23185-2009>

Assessment and benchmarking of terminological resources — General concepts, principles and requirements

1 Scope

This International Standard describes fundamental concepts related to the effective use of terminological data. It provides general principles for a model applicable to a variety of terminological resources. It clarifies the usability attributes that constitute the model and provides guidelines for the overall assessment of terminological resources by taking the user's objectives into account.

2 Terms and definitions

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

2.1

assessment

⟨terminology⟩ process to demonstrate that a **terminological resource** (2.8) fulfils specified requirements

2.2

benchmark

⟨terminology⟩ **usability attribute** (2.11) used as a reference point or metric against which the **usability** (2.10) of a **terminological resource** (2.8) can be measured

2.3

benchmarking

⟨terminology⟩ application of **benchmarks** (2.2) to **terminological resources** (2.8)

2.4

entity

any concrete or abstract thing that exists, did exist, or might exist, including associations among these things

EXAMPLE A person, an object, an event, an idea, a process, etc.

[ISO/IEC 2382-17:1999, 17.02.05]

2.5

model for assessment

⟨terminology⟩ model that identifies the **usability attributes** (2.11) of **terminological resources** (2.8) and their interrelationships

2.6

special language

language used in a subject field and characterized by the use of specific linguistic means of expression

[ISO 1087-1:2000, definition 3.1.3]

2.7

terminological data

data related to concepts or their designations

[ISO 1087-1:2000, definition 3.8.1]

2.8
terminological resource
terminological data resource
entity (2.4) composed of collections of **terminological data** (2.7) with the **usability attributes** (2.11) that are generated by grouping/structuring the data, or incorporating the data into an application

NOTE 1 A terminological resource generally contains **terminological data** (2.7) that are structured (e.g. a terminology database), marked up with a mark-up language (e.g. an XML data file) or associated with a structured layout (e.g. a dictionary). A terminological resource can contain even plain texts (e.g. texts with distinguishable language style) from which usable terminology data can be extracted with the help of modern content-processing technology.

NOTE 2 Generally, terminological resources produced through professional terminological activities have well-organized/structured and high-quality **terminological data** (2.7) and thus have sufficient **usability attributes** (2.11). Those terminological resources created through other processes/activities often have unpredictable quality and usability.

2.9
terminology
set of designations applied to concepts belonging to one **special language** (2.6)

NOTE The definition in ISO 1087-1 is “set of designations belonging to one special language” (ISO 1087-1:2000, definition 3.5.1).

2.10
usability
extent to which an entity can be used to achieve goals effectively, efficiently and satisfactorily

2.11
usability attribute
(terminology) property of **terminological data** (2.7) related to **usability** (2.10)

NOTE 1 Usability attributes can be distinguished as measurable or non-measurable. Measurable usability attributes can be measured quantitatively, whereas non-measurable usability attributes need to be described qualitatively.

NOTE 2 A usability attribute is described according to specified requirements or measured on a given basis.

3 Uses of terminological resources

Generally, a terminological resource is needed:

- to consult the knowledge content related to concepts or concept representations in a special language (e.g. looking up terms in a dictionary, querying a database, searching in a terminology portal);
- to manage domain-specific information (e.g. the management of e-business, e-government, e-learning);
- to exchange terminological data efficiently (e.g. the data exchange between activities, information systems and terminology institutes; importing terminological data to a translation memory; exporting terminological data for compiling a specialized dictionary);
- to facilitate terminological work processes (e.g. works by terminologists);
- to merge distributed paralleling terminological resources for the provision of data services (e.g. on-line services that allow customized data output, cooperative data input or management, semantic web approaches).

4 Terminological resources

4.1 General

The concept of terminological resources shall be understood from the following perspectives.

- a) Terminological data are presented, recorded or stored in data media via processes of data preparation, recording and processing.
- b) Using terminological data involves activities such as
 - 1) processing the terminological data with various technologies to provide a service to users,
 - 2) acquiring proper logic content, epistemological forms and linguistic expressions of specialized knowledge represented by terminological data,
 - 3) transporting/transforming specialized knowledge into the user's intended language context, e.g. translation or localization. However, whether or not a piece of terminological data fits to an intended language context is beyond the scope of this International Standard.
- c) From user's point of view, the two main requirements for usable terminological data are that
 - 1) the terminological data meet the user's need to acquire terminological information content,
 - 2) the terminological data are designed to allow desired access or processing.

Therefore, when terminological resources are analysed, it is generally the case that the terminological data are associated with technological means. In other words, technological means are generally required to make terminological data usable as a terminological resource.

A terminological resource shall be seen as an integrated dynamic system of terminological data. The system begins to exist, to evolve and to function once the terminological data are presented. The usability of the system is demonstrated by its usability attributes. This International Standard clarifies these attributes.

Terminological resources can be systematically assessed. If possible, they may also be automatically assessed. Terminological resources shall be assessed on the basis of analyses of their creation, their management and their potential use.

EXAMPLE 1 One thousand term entries on a digital medium are usually accompanied by a user guide explaining how the medium is accessed and what equipment or tools are required to read the data. The users can then follow the instructions and browse the data to determine if they are useful for their purpose. In this case, the terminological resource is composed of the physical data and the medium that verify the usability attributes explained in the user guide.

EXAMPLE 2 A dictionary of terminologies in mathematics, physics and chemistry is also a terminological resource. In theory, to a user who only needs to access chemical terminology, this dictionary will be little different from a dictionary that has only chemical terminology, i.e. without mathematical or physics-related terminologies. Such a terminological resource will be evaluated differently by users who need to consult mathematical or physics-related terminologies.

These two examples propose that the systematic methodological approach to an assessment focused on the terminological data be

- to take the data together with their overall associations as a terminological resource,
- to assess the usability attributes of the terminological resource through comprehensive analysis of the following elements: data recording, data storage, data format, data structure, the appropriately assigned subject field and user's practical need, etc.

4.2 Model for assessment of terminological resources

A model for assessment of terminological resources shall consist of four sets of usability attributes, each related respectively to

- terminological data,
- data management,
- data output,
- data input.

Not every set of usability attributes is necessarily pertinent to every terminological resource.

EXAMPLE 1 If a hardcopy of a specialized dictionary is considered to be a terminological resource, it will not have significant usability attributes of data input and data management; its usability attributes will only include those derived from the output (printed pages), namely the terminological data that are on those pages and the printed indexes.

Similarly, not all the individual usability attributes related to data management, data output and data input are necessarily pertinent to every terminological resource.

EXAMPLE 2 In the context of a database, an index displayed on user interfaces is not necessary. The user will generally not consider such a displayed index to be a usability attribute.

Figure 1 illustrates the general structure of the model for assessment.

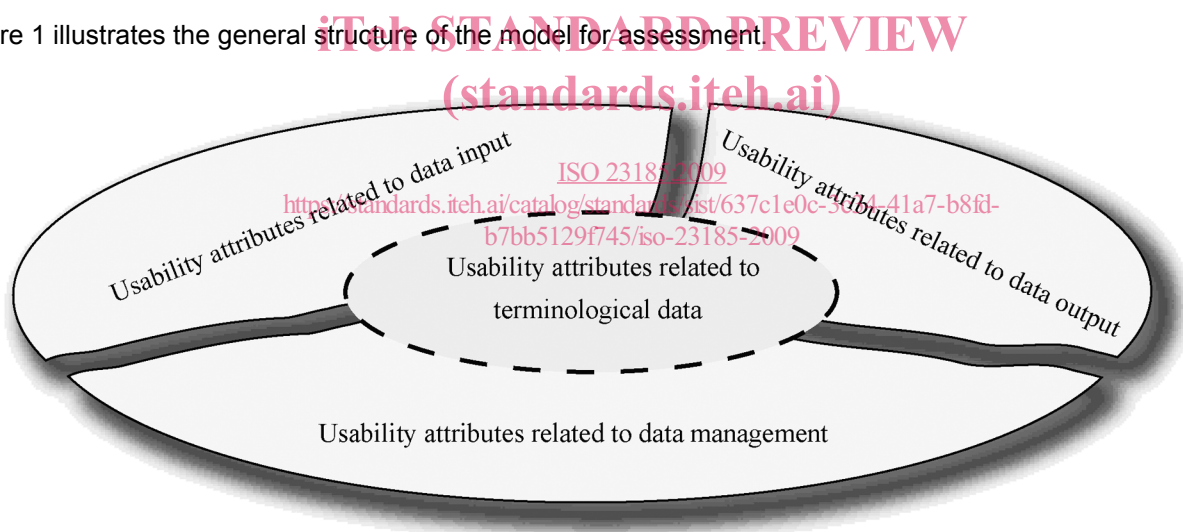


Figure 1 — Four sets of usability attributes

Although in the following clauses the attributes are described individually, it is necessary to bear in mind that the usability attributes can be interrelated with, and dependent on, each other.

4.3 General usability attributes of terminological resources

4.3.1 Usability attributes related to terminological data

A terminological resource shall have enough attributes to meet the complexity of the terminological data while at the same time meeting the requirements for specific purposes, such as

- data structure specification,
- data category coverage,

- subject field coverage,
- language coverage,
- compliance with rules of coherence,
- use of controlled external data,
- use of authoritative sources,
- intellectual property rights ownership indication,
- symmetry of the terminological data collection,
- size of the terminological data collection.

4.3.1.1 Data structure specification

Data structure specification refers to the data model of a terminological resource. With an understanding of the data model, users can employ appropriate technologies to extract terminological data from a terminological resource or to further set up services using a terminological resource, and so on. Therefore, the data structure specification shall be explicit to users. This is the rule for this usability attribute.

4.3.1.2 Data category coverage

Data category coverage refers to the set of data categories that are used in the data model of a terminological resource. The composition of the set of data categories varies depending on subject fields or applications.

EXAMPLE ISO 12616 specifies the necessary data categories for translation-oriented terminography. This provides a metric for determining whether a terminological resource is suitable for translation-oriented terminography.

ISO 12620 has specified possible data categories for recording terminological data.

Data category coverage is measurable if data categories specified in ISO 12620 are used and if a data model is implemented to comply with ISO 16642. If the data model and data categories of a terminological resource do not comply with ISO 16642 and ISO 12620, a mapping mechanism is necessary when evaluating the terminological resource.

4.3.1.3 Subject field coverage

Subject field coverage refers to the set of subject fields in which the knowledge information is represented by the terminological data of a terminological resource.

A terminology deals with special language in a particular field of knowledge. This particular field of knowledge shall be indicated by one or more subject field indicators belonging to

- a discipline in an established classification scheme or controlled language, or
- an application-specific or domain-specific community.

If not, it shall be indicated by explicit subject field indicators. A subject field indicator, either a representation in a controlled language or a commonly used explicit indicator, can refer to a collection of terminological data as a whole, apply to each individual entry, or both.

Subject field coverage is measurable when the controlled languages or explicit subject field indicators employed to indicate subject fields are comparable. Consequently, users can match the subject fields covered by a terminological resource against their expectations.