



15 December 2022

Interinstitutional Data Principles

Data guidelines and data reporting guidelines for IATE

1 Introduction

The Interinstitutional Language Technology Group (ILTG) has laid down *Interinstitutional Data Principles*, which were adopted by the ECT on 12 October 2022 (ECT 22-052 Final) as a step towards an active and common data management policy for the relevant interinstitutional tools, mainly Euramis and IATE. Based on the above-mentioned document, the IATE Management Group has established the following IATE-specific guidelines and related reporting guidelines.

2 Background information on IATE

IATE (Interactive Terminology for Europe) is a dynamic terminology management system designed to support the multilingual drafting of EU texts, and legal texts in particular.

It aims to provide relevant, reliable and easily accessible data which offer distinct added value compared with other sources of lexical information (e.g. translation memories, the internet, electronic archives). Since IATE is the European Union's terminology platform, the entries in it are created and enriched with multilingualism in mind.

IATE has been open to the public since 2007. This, together with the fact that IATE content is used in translation aid software, highlights the importance of managing the data with rigour. Internal users with editing rights in IATE follow specific training and have at their disposal a detailed IATE Handbook which sets out best practices¹ - including copyright rules and how to properly draft references - and describes the various data categories² available.

It is important to emphasise that IATE is the result of the merge of various institution-specific databases, and that the clean-up of this legacy data and consolidation of duplicates has been an ongoing task since the beginning of the project and will continue in the coming years.

¹ Terminology work in the EU institutions and in IATE follows ISO principles and methods of terminology work. More information is available in [ISO 704:2022](#).

² The term 'data category' is used in these guidelines to refer to the specifications of metadata used in language resources, including terminology databases, as per <https://datcatinfo.net/> and ISO standards [ISO 12620-1:2022](#) and [ISO 12620-2:2022](#). As indicated in [ISO 12620-1:2022](#), a data category can be viewed as a generalisation of the notion of a field in a database.

3 Interinstitutional Data Principles applied to IATE

Principle 1: Linguistic quality

The linguistic data stored and managed in IATE must be of a quality that allows its reuse with confidence by all partner services.

The nature of IATE as a terminology management system is to offer users useful, researched, documented and reliable information. The **general quality principles**, listed in the IATE Handbook, are as follows:

A well thought-out IATE entry must give users all the necessary information to allow them to judge whether the proposed solution is appropriate and reliable. It must also allow other terminologists wishing to work on the entry to understand and delimit the concept clearly, and not to confuse it with other related concepts. Each term must have at least one reference that is as reliable as possible, and either a clear definition or a quotation showing how the term is used ('Term in context'), or both.

It is very important to promote the development of IATE entries towards multilingualism, by consolidating and merging any monolingual or bilingual entries for the same concept and by adding as many languages as possible to new entries.

IATE offers **tools that guide users to provide good-quality data**:

- The following fields have been made mandatory, to ensure that a minimum set of added-value information is always available: domain, term, term reference, reliability, term type. Definition reference and context reference are mandatory if a definition and a context are inserted.
- New or modified data enters a validation cycle, in which a terminologist checks the correctness of the content.
- IATE offers enhanced and standardised referencing, with direct querying and retrieval of EUR-Lex titles in the language concerned, URL verification and automatic consultation date, and management of frequently used references via a clipboard.
- IATE contains multiple features adapted to central terminology coordinators' and language terminologists' needs to better manage data (i.e. advanced duplicate detection, advanced search, batch update, etc.).
- Checks are in place to ensure that only technically compliant data are saved in the IATE database.

IATE users are able **to analyse the data quality** using the following features:

- A clear display of data categories in the main IATE search results and data management screens facilitates consultation and decision-taking by users. These categories can also be retrieved programmatically via the IATE search API (application programming interface) and other features' APIs.
- IATE offers numerous filters in the various search, term recognition and data management features, which allow users to identify lower- and higher-quality data.

★ **To ensure that the data in IATE is of high quality**, IATE partners shall commit to adhere to the following:

- Terminology work carried out by IATE partners shall comply with the interinstitutionally agreed guidelines in the IATE Handbook and the best practices set by the various language

communities. Additionally, this can be complemented by internal guidelines adapted to the context of each service.

- Copyright and personal data protection rules shall be adhered to at all times; detailed information can be found in the IATE Handbook.
- All IATE editors (with a translator, terminologist or administrator user role) shall be trained according to their needs, based on the instructions and guidelines explained in the IATE Handbook and language-specific annexes.
- All relevant quality-related instructions shall be included in the IATE Handbook and updated whenever necessary by the IATE Handbook Task Force.
- The best practices shall be promoted in internal communication and training and via the interinstitutional communication platform EurTerm.

Principle 2: Transparency and proper annotation

Data in IATE must be stored and annotated (tagged) in such a way that users can determine its quality, relevance and reliability easily.

IATE offers a very **comprehensive set of data categories** in the three levels of each terminology record (entry - or 'language-independent' - level, language level and term level). These are thoroughly documented in the IATE Handbook. The clearest sign of quality and reliability in the entry level is the **primary star**: it is an indication of an entry's overall quality and preferred status over possible duplicates. Low-quality data can be given '**pre-IATE**' status. Pre-IATE data is not visible to the public, and internal users can easily filter out pre-IATE content from searches and exports. In addition, to give an estimation of the overall completeness of the record, a specific **field completion score** is allocated automatically to the three levels and can be exploited via filters.

Moreover, the system automatically saves useful **administrative information**, which is visible to users and exploitable via filters: created by/creation date; modified by/modification date; validated by/validation date. There is also an **audit system**, which allows users to track changes over time and undo an operation or roll back to a specific point in time.

A collections management module allows entries to be grouped in thematic, project-related or normative subsets, which may be institution- and/or language-specific, interinstitutional or multilingual.

★ When editing data, IATE partners shall **use the available IATE data categories to indicate the quality and reliability** of the content. Important fields for this purpose are the following:

Entry level:

Primary star to indicate a good-quality entry and/or the best entry among duplicates

Entry status 'RAW' to indicate an entry in an early stage of completion or undergoing extensive work

Lifecycle to flag the subject of the entry as 'proposed', 'abandoned' or 'historical'

Management note to share information related to the management of the full entry

Collections to show that the entry belongs to a larger set of data (related to a subject area or a project)

Related projects to show that the entry belongs to a centrally coordinated project

Cross-references to show the relationships between entries

Domains to clearly delineate the subject area to which the concept belongs

Language level:

Definition to delimit the concept clearly

Note to provide any additional relevant information that relates to the concept

References to indicate a credible source or authority in the relevant field

Validation status: either validated or non-validated. In addition, pre-IATE status can be used for data of low quality.

Term level:

Reliability value, on a scale from zero to four stars

Evaluation to show whether the term is preferred, admitted, deprecated, obsolete or proposed

Note to provide any relevant information related to the term

Context to illustrate the use of a term

References to indicate a credible source or authority in the relevant field

Validation status: either validated or non-validated. In addition, pre-IATE status can be used for data of low quality.

Principle 3: Completeness

All partner services shall aim to store and manage all their terminological data in IATE.

Practically all of the IATE partners use IATE as their main terminology repository. European Central Bank (ECB) terminology remains for the time being outside IATE, but it can be accessed separately via Quest, if the user selects the resource. In addition, some language teams maintain local termbases for specific purposes (clients' feedback, phraseology, etc.).

IATE includes import, global update and export features, which facilitate data transfer to and from IATE. It is also possible to manage the workflow of terminology projects within IATE, using the terminology projects module.

★ IATE partners shall endeavour to **store all their relevant terminological data in IATE**, and to do this systematically and in a timely manner, in order to share the results of their work with other IATE partners.

Principle 4: Data sharing and reuse by default

All partner services can reuse the data stored in IATE for any internal purpose.

All IATE partners have the **right to reuse the data stored in IATE** for internal purposes via the IATE user interface, the IATE APIs or data exports in various formats.

Data can be **marked as confidential** (not visible to the public) at entry, language or term level and in different fields (with filters available). It is also possible to **protect** an entry, a language or a term, so that only users or administrators from that institution can modify the data. Detailed rules on protecting content are established and shared in the IATE Handbook.

★ When IATE data are to be exploited in systems hosted outside the EU local infrastructures, the IATE Management Group shall be consulted.

★ **Sharing of internal data** with external parties shall be discussed within the IATE Management Group. **Sharing of public data** with external parties is possible via the IATE public search APIs, with the recommendation to obtain credentials from the IATE Support & Development Team, and the IATE download feature, and is regulated by specific conditions of use set out in the [IATE Legal Notice](#).

Principle 5: Data life cycle management

All partner services shall take appropriate measures to manage the life cycle of their data in IATE.

In IATE, data updates and maintenance are the responsibility of the data owner and fully tracked in the entry history. There is no particular interinstitutional follow-up on this activity, apart from the general follow-up of data modifications on a quarterly basis. Institutions can carry out closer monitoring per entity over a specific period. Joint cleaning efforts are discussed and handled within the IATE Clean-up Task Force. The interinstitutional post-adoption workflow, including the potential automatic identification of obsolete references or dead links, are topics that would need to be investigated further.

IATE users can use the following **features to manage the life cycle** of their data:

- IATE offers various data categories to mark the life cycle of data: **entry status** 'raw' for entries in an early stage of completion, **lifecycle** 'proposed', 'abandoned' or 'historical' for concepts in these stages, **evaluation** for 'proposed' or 'obsolete' terms, **cross-references** for 'predecessor' or 'successor' concepts.
- Content that needs to be reworked can be given **pre-IATE** status.
- The **advanced search** allows users to find data in the above-mentioned categories, as well as data that has not been updated for a certain time.
- With the **batch update** mechanism, IATE administrators can update a large set of IATE entries in one go (to update obsolete URLs or certain references, for example).
- The **batch deletion** feature allows IATE administrators to delete a large set of entries, language levels or term levels with one click.
- A **feedback** mechanism is available in IATE for internal and external users.
- **Bookmarks** or **collections** allow for easy retrieval of specific data subsets.
- A '**watch**' mechanism facilitates the follow-up of specific entries and collections.

★ In order to ensure that the **data in IATE remains relevant and useful**, IATE partners shall commit to the following:

- IATE partners shall take appropriate measures to manage the life cycle of their data in IATE, including timely validation and feedback management.
- Whenever identified, outdated, irrelevant and other low-quality data shall either be given pre-IATE status or deleted.
- IATE partners shall invest in the clean-up of legacy data and consolidation of duplicates, and discuss the joint cleaning efforts within the IATE Clean-up Task Force.
- A thorough discussion on managing the life cycle of IATE data shall be held at interinstitutional level, and the IATE development team shall investigate the possibility of developing suitable IATE features to further facilitate data life cycle management.

4 Data reporting guidelines applied to IATE

The IATE Management Group shall promote these data management guidelines and adapt them to evolving needs and technical possibilities, in consultation with relevant task forces and groups. It will report on the compliance with these principles in its annual activity report.

The report will contain a **briefing on the data-management-related activities** of the following task forces and groups:

- IATE Management Group
- IATE 2 Task Force
- Clean-up Task Force
- Central terminology coordinators
- IATE Handbook Task Force

When reporting on the data management guidelines, the following **qualitative and quantitative indicators** will be used. These indicators are monitored on a quarterly basis and they are available in the interinstitutional collaborative platform [EurTerm](#) ('Quarterly statistics' section):

Global figures

- Total number of entries
- Total number of terms

Qualitative indicators

- Multilingual entries (>10 languages)
- Duplication level
- Primary entries
- IATE entries with added value
- Validation rate for terms

Quantitative indicators

- Entry levels created, modified or deleted
- Language levels created, modified, validated or deleted
- Term levels created, modified, validated or deleted

Other potentially useful indicators shall be considered by the IATE Management Group and incorporated in the reporting as needed.