Project acronym:	**EUROTERMBANK**		
Project title:	**Collection of Pan-European Terminology Resources through Cooperation of Terminology Institutions**		
Contract number:	**EDC - 22 267 - EUROTERMBANK**		
Starting date:	**01/01/2005**	**Ending date:**	**31/12/2006**

D1.2 *FINAL METHODOLOGY REPORT*

Version: *V 1.1*
Date: *31/12/2005*

Dissemination: *Public*

Contributors:

Sabiedrība ar ierobežotu atbildību Tilde
VIENIBAS GATVE 75A, 1004, RIGA,
(Tilde, Latvia)

The Institute for Information Management,
Fachhochschule Köln (University of Applied Sciences Cologne)
CLAUDIUS STRASSE 1, 50678, KOELN,
(IIM, Germany)

Center for Sprogteknologi, University of Copenhagen
NORREGADE 10, 1017, COPENHAGEN K,
(CST, Denmark)

Lietuvių kalbos institutas (The Institute of the Lithuanian Language)
P. VILEIŠIO G. 5, 10308, VILNIUS,
(LKI, Lithuania)

Terminology Commission of Latvian Academy of Science
AKADEMIJAS LAUKUMS 1, 1050, RIGA,
(LAS, Latvia)

MorphoLogic Számítástechnikai Korlátolt Felelősségű Társaság
ORBÁNHEGYI ÚT 5., 1126, BUDAPEST,
(MorphoLogic, Hungary)

Department of Estonian and Finno-Ugric Linguistics, Faculty of Philosophy,
Tartu Ülikool (University of Tartu)
ÜLIKOOLI 18, 50090, TARTÜ,
(TÜ, Estonia)

OŚRODEK PRZETWARZANIA INFORMACJI (Information Processing Centre)
AL. NIEPODLEGŁOŚCI 188B, 00-608, WARSAW,
(OPI, Poland)

Authors:
Albina Auksoriūtė, Institute of Lithuanian Language (LKI), Albinaa@lki.lt
Audronė Bielevičienė, Institute of Lithuanian Language (LKI), audrone@gironda.lt
Tomasz Borkowski, OPI
Juris Borzovs, Latvian Academy of Sciences, juris.borzovs@riti.lv
Anna Braasch, Centre for Language Technology, anna@cst.dk
Eduards Cauna, Latvian Academy of Sciences, cauna@parks.lv
Anita Dravniece, Latvian Academy of Sciences, int@lza.lv
Nijole Dudlauskiene, Lithuanian Standards Board, nijole.dudlauskiene@lsd.lt
Elzbieta Gorzkos, OPI, egorzkos@opi.org.pl
Lina Henriksen, Centre for Language Technology, lina@cst.dk

Heiki-Jaan Kaalep, Tartu University, hkaalep@psych.ut.ee
Danuta Kierzkowska, TEPIS Publishing House, dk@polishlaw.com.pl
Balazs Kis, Morphologic, kis@morphologic.hu
Bente Maegaard, Centre for Language Technology, bente@cst.dk
Claus Povlsen, Centre for Language Technology, claus@cst.dk
Marin Raguz, Institute for Information Management, marin.raguz@fh-koeln.de
Inke Raupach, Institute for Information Management, inke.raupach@fh-koeln.de
Klaus-Dirk Schmitz, Institute for Information Management, klaus.schmitz@fh-koeln.de
Krystyna Siwek, OPI, ksiwek@opi.org.pl
Valentīna Skujiņa, Latvian Academy of Sciences, vaska@lza.lv
Robertas Stunzinas, Institute of Lithuanian Language (LKI), robertasst@lki.lt
Andrejs Vasiļjevs, Tilde, andrejs@tilde.lv
Uldis Priede, Tilde, uldis.priede@tilde.lv
Ainars Blaudums, Tilde, ainars.blaudums@tilde.lv
Raivis Skadiņš, Tilde, raivis.skadins@tilde.lv

Editors:
Centre for Language Technology, University of Copenhagen

CONTENTS

1 General introduction

This report provides recommendations for best practice within the field of terminology work. The report is based on D.1.1 *Current Standards and Best Practices Assessment Report* which describes relevant terminology standards and current terminology processes in selected terminology resources, bodies and projects belonging to the new EU member countries as well as the old EU member countries. D.1.1 thus established a sort of baseline or map of current terminology methodologies which have been analyzed and elaborated in order to extract best practice within all the aspects of terminology methodologies.

A terminology process comprises a complex myriad of tasks, procedures and principles. In order to establish a solid structure and provide a reasonable overview, the report groups tasks and principles in suitable blocks constituting the chapters of this report, for example *Workflow of terminology work, Concept analysis, Exchange format* etc. (the structure of this report is broadly the same as chapter 7 of D.1.1).

1.1 Best practice in different scenarios

Best practice of terminology work is not exactly the same in all contexts and therefore it has been necessary to establish some scenarios to represent schematic frameworks of terminology work. Three different frameworks of terminology work are represented in D.1.1: the organizational, the national and the international level and the description of best practice in this report is based on this distinction.

In order to analyze and assess methodologies applied in each level it has been necessary to identify the conditions and criteria that are especially characteristic of each level. Conditions mean the state of things that cannot (or hardly can) be changed. For example a condition might be that all language professionals of a terminology team do not have access to the internet. This may of course have an impact on the execution of particular routines. Criteria mean the objectives or aims of the terminology work. For example a criterion might be that newly created terminology should be available to other language professionals as fast as possible (even at the expense of quality checking). For further information about collection of assessment data please see Appendix A.

Even within one level conditions and criteria may differ widely depending on the nature and size of the particular organization, the purpose of the terminology: for example internal or external use, knowledge sharing or translation purposes etc. Therefore best practice described for one level may also be applicable for an organization that belongs in another level. For example, best practice described for the international level may be the best solution for a private company (belonging to the organizational level) with heavy translation requirements.

The scope of the organizational, national and international frameworks are defined below together with descriptions of conditions and criteria.

1.2 Introduction to organizational level

The two main types of activities that take place at the organizational level and involve terminology work are, besides translation and creation of documents, to create new term bases or terminology dictionaries.

An organization at the organizational level is, in this context, to be understood as a group of people who are working together on terminology tasks (e.g. two translators, one terminologist, one language specialist, and several domain experts).

A characteristic feature is that terminology created at organizational level, as a rule, is limited to only one domain (or several closely related domains) and harmonization does therefore not play a significant role. Another characteristic is that speed usually is an important criterion.

Examples of such organizations would be: private companies (such as translation bureaus), national ISO bodies, EU translation offices, ministries, and faculties of universities.

As mentioned above the type of organization sets the scene in terms of basic conditions and high priority criteria in the terminology work. The tables below provide a walk-through of the conditions and criteria that apply at the organizational level:

Existing conditions	**General comments**
Access to terminology tools/ no access to terminology tools	At the organizational level terminology developers will in some organizations have access to terminology tools - in other organizations terminology tools will not be available. The descriptions of best practice will focus on the situation where no terminology tools are available and refer to methodology described at the other levels if terminology tools are available.
At least translators and domain experts are represented	Terminologists are often not part of the terminology developer team..
Focused domain coverage	At the organizational level the terminology developer team usually only works within a very limited number of domains. Therefore, harmonization issues in general are not relevant to touch upon in this context.
Size	At this level, organizations are usually rather small. In this context the number of language professionals is defined to be 2-10.

Criteria with a high priority	**General comments**
Speed and up-to-dateness	At this level speed of terminology production is usually an important criterion. The financial situation

	will often demand that time is an important factor and it is often required that new terminology is available to other users as fast as possible.
Quality (consistency, unambiguity, feedback, public acceptance)	Quality is per definition an important criterion. At this level, however focus will be on best practice where speed and quality coexist. That is: how can an organization perform terminology work of an acceptable quality within a limited timeframe.
Design reflects many types of user requirements / Design: easy to use and maintain	

Note: Exchangeability has not been mentioned as a criterion at this level as it is often not a factor taken into consideration. It is however still recommended that even at the organizational level database and terminology developers take this criterion into consideration when developing a new database for terminology.

1.3 Introduction to national level

In broad terms main activities at the national level are to coordinate and regulate the terminology work within the territory of a state. More specifically this overall task includes management of:

- approval/endorsement of terms created by sub-organizations
- harmonization of these terms with existing terminology classified as belonging to other subject domains and/or concept systems

As a consequence of these areas of responsibility many sub-organizations and many domains are involved in this terminology work.

At the national level, a terminologist or a group of terminologists, permanently provided with consultation of domain experts, are responsible for making relevant decisions.

Additional features characterizing terminology work at the national level would be that the terminology coordinated is usually monolingual or bilingual and that the time and financial conditions (although the state support is sometimes inadequate) are not as crucial as at organizational level.

The tables below show the conditions and high priority criteria to consider at the national level.

Existing conditions	**General comments**
Access to terminology tools	At this level we presume that language professionals have access to terminology tools.
At least (translators) terminologists and domain	At this level we presume that all types of language professionals are available.

experts are represented	
Broad domain coverage	Language professionals work within many domains.
Satisfactory financial situation	
Involvement of relevant institutions / terminology coordination	
Usually monolingual or bilingual	Terminology collections are usually mono- or bilingual. Terminological work in a multilingual environment is treated at the international level.

Criteria with a high priority	**General comments**
High quality in general terms	In this context 'high quality terminology' means that terminology development is based on sound research principles.
Exchangeability	Exchangeability means exchange of terminology using ISO standard approved exchange methodology.
Availability	Terminology should be available to external users.
Design should reflect many types of requirements / Design: easy to use and maintain	
Harmonization	Harmonization of terminology is important and is probably an inherent part of the quality-criterion.
Design should reflect many types of requirements / Design: easy to use and maintain	

1.4 Introduction to international level

Similar to the national level, the primary activity at this level is concerned with coordination and management of terminology work in a well-organized infrastructure. Consequently, terminology work at the international level concerns approval/endorsement of already coined terms as well as harmonization of these terms so that they will be in compliance with the already existing data in the term bases.

The main distinctive element separating the national from the international level is the number of languages to be treated in the terminology work. While the national level usually is either mono- or bilingual in terms of coverage, the international level always spans over several languages.

The tables below describe in more detail the framework, i.e. the conditions and high priority criteria which are relevant seen from an international environment.

Existing conditions	**General comments**
Access to terminology tools	
At least (translators), terminologists and domain experts are represented	It is presumed that all types of language professionals are available
Broad domain coverage	Language professionals work within many domains
Satisfactory financial situation	
Involvement of relevant institutions/terminology coordination	
Multilingual	It is presumed that terminology collections are usually multilingual

Criteria with a high priority	**General comments**
High quality in general terms	Interpretation of the concept of 'high quality terminology' is that the work is based on sound research principles.
Exchangeability	Exchangeability means exchange of terminology using ISO standard approved exchange methodology.
Availability	Terminology should be available to users outside the international cooperation.
Design should reflect many types of requirements/design: easy to use and maintain	
Harmonization	Harmonization of terminology is important and is probably an inherent part of the quality-criterion.

2 Overall workflow of terminology tasks

Terminology work is performed differently in different countries. It depends on the organizations and institutions engaged in this work, as each institution defines to a certain extent different goals and objectives. The way terminology work is arranged depends on the specific goals and activities as well as on who is involved in the management and development of the terms. The stages of terminology work differ as well: some institutions have few stages for terminology development; others include several stages into the process.

These main aspects of terminology work will in the following be summarized in terms of organizational, national and international levels.

2.1 Organizational level

2.1.1 Types of organizations

The new EU member states have different organizations engaged in terminology work and they of course organize the terminology work differently and apply different work practices.

The following main types of organizations have been identified:

1) Translation agencies, for example: Estonian Legal Language Centre (Eastern), Translation and Terminology Centre in Latvia (TTC), Translation, Documentation and Information Centre in Lithuania and others.
2) National ISO bodies, for example: LSB in Lithuania, PKN in Poland, LS in Latvia and others.
3) Research institutions, for example: University of Tartu, Institute of the Estonian Language, Institute of the Latvian Language, Terminology Centre at the Institute of the Lithuanian Language and others.

EU translation offices, terminology committees under the Ministries and others could be listed here as well.

2.1.2 Workflow

Organizations translating EU legislation. First, translators search for the corresponding terms needed in their national language in related documents, terminology glossaries and available terminology databases, and also in relevant literature of the domain in question. If they fail to find a corresponding term in their national language, they develop new terms or borrow them from other languages. In connection with creation of new terms it is important to get as close as possible to the main and the specific features of the concept being defined. Furthermore, all newly coined terms have to be correct, consistent and comply with the rules of the national language (e.g. as regards compounding).

National ISO bodies develop national versions of the European or international terminology standards, adapt them or develop national terminology standards. Developers

of standards are usually specialists of a certain domain, thus they propose terms that are already used in their domain of activities. When a term does not exist in the national language, they develop or borrow a new term

Research institutions deal mainly with theoretical research on terminology, but terminologists at these institutions also do practical work on normalisation of terminology and work with specialists of various domains as language experts. Terminologists give recommendations to specialists of a particular domain re. naming the concepts in the most appropriate way.

2.1.3 Differences and common features

The organizations differ not only in terms of the type of their work but also in terms of the staff employed. Translation centres usually employ translators, language specialists, sometimes experts of different domains are involved as well, while the involvement of terminologists is quite rare. The terminology work at the national ISO bodies is usually entrusted to the specialists and experts of different domains, however terminologists are not always involved. The activities of research institutions usually involve only linguist terminologists. Only in connection with general terminology projects will relevant specialists and terminologists be invited in their capacities of consultants or experts.

The entities working at an organizational level have one central feature in common: they manage and develop terminology of one or few, usually related, domains. Therefore, harmonization of concepts and terms between domains is usually not possible. This is one of the main differences between the organizational level and the national and international levels.

2.1.4 Existing conditions and recommendations

Terminology sources and tools. The types of terminology sources and tools available to the terminology developers are very important. A source available to everybody is terminology dictionaries. The new European member states have quite a number of issued paper terminology dictionaries in different domains; electronic terminology dictionaries gradually emerge as well.

Differences between levels. Terminologists are rarely involved at the organizational level (as opposed to the other levels) which means that the translators' qualifications, professional background, knowledge of the native language and main principles of terminology are of crucial importance.

Terminology developers at the organizational level do often not have access to the international terminology banks and internet-based databases, whereas a similar problem is not faced on the national and international levels.

Speed. Speed is a basic requirement in terminology management and development tasks on the organizational level. The translation bureaus have to observe the deadlines for translation of different documents and the national ISO bodies have to adhere to the deadlines for preparation of e.g. the national versions of the international standards.

Approval. The terminology developed on the organizational level, as a rule, is not approved and endorsed by any competent national organization, while such practice is common to the national level. Nevertheless, some organizations seek to have their terms approved, e.g. the Lithuanian Standards Board submits all its terms to the State Commission of the Lithuanian Language for approval. In this respect the practice of LSB may be considered an example of good practice.

Quality. A further relevant issue is the quality of terms developed and introduced by the organizations. At the national level, the terminology management and development process involves several stages, and here consistent, correct and adequate terms are introduced, which are harmonized with the terminology in other domains too. Although such a process is quite time-consuming, a comprehensive terminology development process is not only employed at the national level, but it can be observed at the organizational level as well.

Recommendations. It is very important to perform terminology management and development tasks at the organizational level effectively, and to ensure at the same time reliable and high-quality terminology. This objective can be achieved only by involving high-profile translators and different specialists with thorough knowledge of both the language and the key principles of terminology work. The staff should be able to consult reliable available terminology sources. In addition to that, the work quality significantly changes, where the staff includes terminologists or where terminologists are being constantly consulted. Quite a few entities on the organizational level apply such practice.

2.2 National level

2.2.1 Types of organizations

At the national level, as well as at the organizational level, organizations work in various ways. Terminology work at the national level is done in several institutions and organizations:

- Terminology Council of the Hungarian Language
- Terminology Commission of the Latvian Academy of Sciences
- Council for the Polish Language
- State Commission of the Lithuanian Language
- And others.

2.2.2 Workflow

In spite of the differences in activities, the basic tasks are common to most institutions and organizations as follows:

- Terminology and language planning
- Development of integrated terminology systems concerned with international principles
- Standardization and approval of terms
- Maintaining terminography

- Coordination of terminological work in state institutions, standardization departments, translation centres and other organizations.

In some organizations, the activities carried out are broader and comprises not only terminology. For example, the main tasks of the State Commission of the Lithuanian Language are normalization and standardization of the language, implementation of the official language status, coordination of the language policy and terminology work.

Terminology work consists of two stages: expertizing and approbation. Extraction and creation of terms and definitions are carried out at the organizational level (cf. the description of the organizational level), expertizing on and approbation of terms and definitions are carried out at the national level.

At the national level the most important task is the harmonization between domains, because many experts from different institutions take part in the terminology creation workflow, such as domain experts and high-level language specialists (terminologists). For example, at present the Terminology Commission of the Latvian Academy of Sciences consists of 26 subject field sub-commissions. All these work according to the general principles and methodology for terminology work, which are accepted by TC of LAS. Terminologists usually specialize in some particular fields of science, and every field of science has its own specifics. Thus, a terminologist has not only to be familiar with his particular field, but also co-operate with specialists for creating a systematic terminology of that field.

2.2.3 Existing conditions and recommendations

In terminology work, the access to terminology tools and to the internet is very important. The use of terminology tools supports an efficient cooperation between domain specialists. Term banks function as the main terminology tool of actors in terminology workflow. The conception, workflow and methodology of such banks are mostly defined legally. For example, the purpose of Term Bank of Lithuania is to ensure a consistent usage of normalized Lithuanian terms, especially in legislative documents, to create a common information system for various institutions with the possibility for other persons and legal entities to get connected to it, etc.

The time and financial conditions (although the state support is sometimes inadequate) at the national level are not as important as in organizational level. Terminological work at the national level is mainly thorough and qualified terminological development.

At the national level terminology collections are being created for national-wide spread and usually are monolingual or bilingual.

2.3 International level

2.3.1 Types of organizations

At the international level, as well as at other levels, organizations work in different ways. Examples of terminological work at the international level:

- Production of comprehensive, high-quality and reliable terminology stored in term bases (e.g. IATE).
- Terminological publication (e.g. ISTO).
- Terminological standardization:
  - Standardization of principles, methods and applications related to terminology (ISO TC 37),
  - Standardization of terms (e. g. ISO).
- Others.

2.3.2 Workflow

Workflow of these different types of terminological work is based on well-defined tasks and procedures. For example, terminological standardization at the international level is regulated strictly (workflow is based on its own rules). One of best-known international organizations working with terminology standardization is ISO. Its committee, TC 37, develops terminology standards by a six step process:

1. proposal stage
2. preparatory stage
3. committee stage
4. enquire stage
5. approval stage
6. publication stage

2.3.3 Existing conditions, common features and recommendations

Although the character of terminology work in these mentioned types of organizations differ, it is possible to identify common typical features and provide some general recommendations. First of all, terminology collections at the international level are multilingual. That is a very important point not typical to other levels and multilingualism requires language professionals of different languages. Furthermore, one common database is created for all participating institutions, organizations etc. (e.g. IATE) and much attention should be paid to exchangeability. Terminology tools should be integrated into translation and office automation environment.

Actors of terminological work are domain experts, different types of language professionals (translators, interpreters, expert translators, terminologists) and others. Thus terminology harmonization between domains is an essential part of the terminology work.

Actors in a terminology workflow should have access to terminology tools and Internet and they should be able to communicate on-line.

High quality in general terms can be achieved by means of developing appropriate procedures, involving highly qualified domain experts and terminologists (from various countries) and employing strict validation procedures. Well-prepared procedures in the terminological workflow ensure the achievement of the overall high-quality goal.

Terminology coordination is a very important aspect of terminology work at the international level. Various institutions from different countries are involved in the international terminology work and therefore terminology work should be coordinated between countries and institutions.

It is recommended that terminology is available to users outside the international cooperation through on-line access, either free of charge or for a fee. ISO terminology, for example, is only available for a fee.

3 Classification systems

Classification systems are used to organize the terms of a term collection which implies that:

- A classification system should cover all subject fields (domains), in which terminology work is done
- Concepts (and terms) are examined in relation to a subject field, thus one of the most important tasks is to understand and define exactly what a subject field covers
- A classification system helps to understand concept related with term
- A classification system helps to facilitate retrieval of information from term bases.

There are several classification systems used in a terminology context to describe subject fields. Two of the most popular are *Lenoch* and *Eurovoc*.

Lenoch classification system (*http://www2.uibk.ac.at/translation/termlogy/lenoch.html*) is developed by Institut für Translationswissenschaft (Universität Innsbruck). This system is widely adopted by European Union terminology organizations (e. g. EURODICAUTOM).

Eurovoc Thesaurus (*http://europa.eu.int/celex/eurovoc*) is a multilingual thesaurus covering the fields in which the European Communities are active. It provides a means of indexing the documents in the documentation systems of the European institutions and of their users. The European Parliament, the Office for Official Publications of the European Communities, the national and regional parliaments in Europe, some national government departments and European organizations are currently using this controlled vocabulary. Eurovoc Thesaurus can be used also as a terminology classification system.

ETB partner	Classification system
Estonia	Lenoch
Latvia	Lenoch
Lithuania	modified Eurovoc
Hungary	Local
Poland	Local

If we compare both systems it is evident that Lenoch is more straightforward. On the other hand, the Eurovoc system is more *systematic* with logical hierarchy instead of Lenoch “top of the pops” (more popular appear at top level) policy.

Important feature of Eurovoc — it is a *controlled* vocabulary with official translation in at least 20 languages. It is also more up-to-date and better maintained.

Both systems cover subject fields used in everyday terminology practice. Eurovoc could be considered as a better choice owing to ongoing support and frequent updates, as well as availability in more than 20 languages.

3.1 Organizational level

At organizational level mostly terms from one or related subject fields are prepared. It is not that important which classification system is used. However taking into account that terms later will be used in national context, it is advisable to distinguish at early stage exact domain in which term was used. Also it is advisable use the same classification system used for national coordination.

Best practice is usage of Lenoch or Eurovoc at this stage. Eurovoc could be considered as better choice due ongoing support and frequent updates, as well as availability in more than 20 languages.

3.2 National level

At national level it is important to use advanced terminology classification system. It helps to distinguish between similar terms and similar concepts and make decisions whether one term can be used in several domains (e.g. *monitoring* in economy, administration, customs, and environment protection). It also helps to create better translations (indicating field of applicability) and many other advantages. It is important to remind that by very nature terminology is domain-oriented and only nowadays domains are overlapping at accelerated rate.

Best practice is usage of internationally recognized terminology classification system (Lenoch or Eurovoc) at this stage. Eurovoc could be considered as a better choice due to ongoing support and frequent updates, as well as availability in more than 20 languages.

3.3 International level

At international level it is mandatory to use some advanced and internationally recognized terminology classification system.

It is evident that Lenoch or Eurovoc can be used (examples of similar environments are EURODICAUTOM and IATE).

4 Source identification

The expression "source identification" has at least two meanings:

1. Tracing material that contains evidence about the terms in question – their meaning, usage, translations etc. Terms may be found in various types of written material (books, internet), or in oral communication.
2. Registration of referential information that is necessary for finding and consulting the resource again, for estimating the reliability of the resource etc.

When we look at "source identification" in the first sense, we notice that different sources are not regarded as equal. Official documents are usually treated as more trustworthy than unofficial ones; written sources more trustworthy than oral ones. The exact criteria for ordering the sources according to their reliability, however, may be different for different organizations. For an individual organization, an in-house database may be the most authoritative source, because the consistency of the in-house terminology may be more important for the organization than official standards.

A short list of recommendations for preferences comprises the following:
1. A written source is preferable to an oral source
2. A newer source is preferable to an older source
3. An officially published source is preferable to an unofficial source
4. A specialized source is preferable to a non-specialized source (e.g. an LSP dictionary is better than a general language dictionary; a specialist journal is better than a daily newspaper)
5. An original source is preferable to a translated source.

Notice that the above recommendations may result in conflicting solutions. For example, when having to choose between an oral advice from a native expert, and a translated article from a newspaper, the choice will be different, depending on which criterion one regards the more important.

Notice also that there are no recommendations about what is to be preferred, a small or a big dictionary, a freely accessible database or a standard which is available for fee only, a dictionary or an article. Clearly, some sources are easier to use, and some are more widely spread. These aspects, however, are not directly connected with the quality of the resources.

When considering "source identification" in the second sense, it may seem that good practice would be to record as many referencing details about the source as possible, e.g. the exact page of a book where a term is used, or the exact time when the term was first proposed by an expert. However, this would involve spending time and money, and in some cases the outcome might be of little value: for example, when the terminology of a field is not yet established, or when the terminology work is only a by-product of the translation of a text, involving many ad hoc coinages of new terms.

A general recommendation for "source identification" in the second sense would be to record at least the bibliographical information of the source (author, title, publication year). In case of an Internet source, the exact address and date of consulting the address are needed, and in case of oral communication the informant's name, title, affiliation, and date of communication are relevant. If possible, that is, if time and resources permit, it would be advisable to record key words, hyperlinks, ISBN numbers, informant's contact information and other details, which make the identification and consulting of the source easier.

4.1 Organizational level

4.1.1 "Source identification" as tracing relevant material

There are two main types of activities taking place at the organizational level, which form the basis in creation, dissemination and harmonisation of terminology:
1. Translating and creating documents
2. Creating new term bases or terminological dictionaries

These activity types result in different best practices, inevitably.
When the main task of an organization is to translate documents, or to create new ones, the organization should try to use the established terminology as much as possible and avoid coining new terms. The latter may be characterised as a bad practice, resulting from the low qualification of the translators or writers, or from the time pressure, preventing them from spending time on searching for the best solution. However, it may also happen that coining a new term is inevitable: there is no established term, or the expected audience of the text requires it. For example, a user manual, or an advertisement leaflet may need to use a language different from the language used by specialists of the field. Still, this is not a very frequent a situation.

The best practice would be to follow the principles below:

1. Use written, well established, authoritative sources, even if they are in conflict with each other.
2. If there is more than one authoritative source, use the newer one.
3. If possible, avoid creating new terms.
4. If you discover that you have used a wrong term in previously published documents, try to correct this in subsequent editions, and not continue using the same term in subsequent documents. That is, the term repository of the organization itself needs not be considered a definitive, "well established, authoritative" source.

When the main task of an organization is to create new term bases or dictionaries, it is inevitable that new terms have to be found and described, and existing sources used critically. We assume here that the organization hosts skilled language professionals. The main emphasis should be on reflecting the actual usage of professional language. Harmonising is considered a good practice, while term creation is not considered a satisfying solution.

5. The original written and oral communication with the specialists of the field is the most authoritative source of the terms. If they use a term that is different from the ones found in written authoritative sources, then in this particular case, the authoritative sources should not be trusted.
6. A newer source is preferable to an older one.
7. It is not important whether a source is official or not.
If possible, avoid creating new terminology.

4.1.2 "Source identification" as referential information

In published texts (originals or translations), the source of terms is seldom indicated. The same is usually true for term bases or dictionaries, where the sources are listed, following the bibliographic referencing conventions, only once for the entire term collection.
Thus the "source identification" as referential information is mostly used in an in-house term base. The principles, outlined in the general introduction, apply in the organizational level.

4.2 National level

4.2.1 "Source identification" as tracing relevant material

The national level operates on the results, that are available from the organizational level. The national level is preoccupied with harmonising, and not with creating or using terminology (in texts). Thus the best practice for the national level is to use only written term collections as sources. Term collections, intended for the specialist user, should be regarded as more trustworthy than collections, intended for the wider audience (e.g. a term base of laser physics is more trustworthy than a school dictionary of physics terms). It is best practice to prefer newer sources to older ones (with a critical eye on them, though).

4.2.2 "Source identification" as referential information

In a harmonisation process, it is best practice to have as much referential information at one's disposal as possible, as this may help in making well-founded decisions. In addition to what was mentioned in the introduction, one may need information about the professional background of the author(s), about the time when the term collection was created (as opposed to the time it was published), and about the aim of creating the term collection (e.g. a glossary, supplementing a PhD thesis, or an in-house term collection).

4.3 International level

The international level is similar to the national level in that it operates on the results that are available from the organizational level. If a term collection has been harmonized at the national level, it can be viewed as a result from the organizational level with some extra parameters like high quality. Thus the best practice at this level is not different from the best practice applied at the national level.

5 Use of the internet as a resource

The Internet as a means of terminology work has two very different purposes:

(1) A large but unstructured set of textual and database resources to be searched when terminology data are being created.
(2) A means of communication and for creating distributed working environments.

In this context focus is on definition (1), but definition (2) could also be considered. The latter being the means of establishing a future network of reliable terminology resources.

Survey
Here follows a brief summary of the contributions of ETB partners, assessing the importance of the Internet as a terminology resource, regarding the various quality criteria set forth earlier.

The table below contains only those rows that were mentioned as important to some extent by at least one partner. Countries where numeric assessment was unavailable, are not listed.

Parameters/criteria	EE	HU	LT	PL
(I) quality in general terms	3		1	
(IV) Possession of auxiliary tools in the terminological work		2	1	
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists		2		
(VI) Establishment of (well-prepared) procedures in the terminological workflow		3		
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality		2		

Assessment of the importance of the Internet as a terminology resource at national organizations in different countries

Estonia	Internet as a term resource is highly unreliable, so there are strict in-house criteria for differentiating between trustworthy and untrustworthy sources.
Hungary	In the process of creating the EUJog termbank, the Internet was used only to access the EU's own resources and to provide a means of communication between the participants of the project. However, for non-formal terms (i.e. those other than names of institutions, positions etc.), search engines were also used as research tools.
Latvia	Internet is used for the following purposes: 1. Search for the existing terms in national language (www.termnet.lv, completedb.ttc.lv, www.termini.lv) 2. Search for the definition of new (unknown) terms (mostly Google "define:term" function, also specific domain glossaries) 3. Search for the usage examples of new (unknown) terms (mostly Google) 4. Search for the translation of the terms or term parts (www.onelook.com, lingvo.yandex.ru, www.multitran.ru) 5. Search for the etymology of the terms or term parts (www.etymonline.com) 6. Search for translations in other languages. Internet is a convenient resource for terminology information (origin, meaning, usage) in most fields. However, it takes time to find all information before creation of a national term and quality vs. time factor appears. More or less full analysis is done for approximately 10% of new terms.
Lithuania	The Internet is not used often as a resource (1). There are no auxiliary tools for this work (1)
Poland	Google is the most popular search engine in Poland, used whenever needed.

Textual assessment of the Internet as a resource by contributing partners

Local organizations were much less featured in this survey. Input was provided by Estonia and Hungary where terminology work is apparently less standardized or centralized than in other partner countries.

Parameters/criteria	EE	HU
(I) quality in general terms		2
(IV) Possession of auxiliary tools in the terminological work		2
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists		3
(X) Thorough validation routine → high reliability		2

Assessment of the Internet as a resource by partners

Estonia	Internet as a term resource is not usable in this context.
Hungary	Interned is used as a research tool and also a means of publishing the termbase. Translators access the termbase using their CAT tool (not a web browser!) over the Internet.

Textual assessment of the Internet as a resource by partner countries

5.1 Recommendations

The Internet as a source of information

The Internet, as defined in (1), is a useful research tool in the process of building terminology entries. However, it cannot be treated as one single coherent resource. There are numerous resources at various levels of reliability that should be systematically classified before using in terminology work.

Reliability classes of Internet-based resources can be listed as follows:

(1) Authoritative resources such as standardised terminology databases (highest reliability)
(2) Private resources of large non-governmental organizations (company-wide terminology databases, or non-standardized terminology databases of international organizations)
(3) Community-built resources (such as Wikipedia)
(4) Other private resources (such as Web-based dictionaries maintained by companies or natural personae)
(5) Scattered textual resources accessible through general search engines (e.g. Google; lowest reliability)

When using the Internet for formalized terminology work, each resource must be assigned to a reliability class, and used accordingly.

The Internet as a source of information can be used for the following purposes, as extensively listed by the Latvian Academy of Sciences:

1. Search for the existing terms in national language
2. Search for the definition of new (unknown) terms
3. Search for the usage examples of new (unknown) terms
4. Search for the translation of the terms or term parts
5. Search for the etymology of the terms or term parts.

The Internet as a means of communication

The Internet as a technology is very important in teamwork in general. As regards EuroTermBank, it is recommended to establish and maintain a standardized way of using the Internet for terminology access. This means the creation and publishing of interfaces that can be implemented by terminology providers (compare with requirements in the user requirements analysis, covered by deliverable D3.1).

6 Term extraction

In a term extraction process a corpus or other collection of texts is systematically scanned for terms. The delimitation of the subject field (domain) is highly dependent on the theoretical background and the practical motivation (e.g. users' request) of the work.

The aim of the process is to record and structure relevant information found in the text sources, this can be regarded as the linguistic dimension of terminology work with the following main tasks to be dealt with:

- Identification of concepts belonging to the subject field, and registration of their designations (terms) and definitions
- Observation of typical linguistic contexts and term usages.

Mostly two strategies are used:

1. Term extraction from bi-lingual parallel text corpus (e.g. "PolTerm translation memory")
2. Term extraction from monolingual source text (e.g. MorphoLogic–Kilgray term extraction tool).

Both strategies can be realized manually or automatically. In ETB partner countries methods are used as follows:

ETB partner	Terms extraction method
Estonia	Manually
Latvia	Manually
Lithuania	Manually
Hungary	Automatic
Poland	Semi-automatic

Best practice
Term extraction is a complementary device to ease terminologists' work. However, it is advisable to analyze text and automatically create frequency-based term lists to avoid omission of most frequently used terms and collocations. Once recognized, these terms and collocations should be translated uniformly throughout all the text.

Best practice includes both automatic and manual extraction. Automatic extraction is most productive, but human interaction is often necessary to make context and knowledge-based decisions. Such a semiautomatic approach is used in PolTerm. The approach involves checking candidate terms against both the PolTerm term database and the PolTerm translation memory (TM). The term is considered new when neither the PolTerm term database nor the PolTerm TM, which together form the so-called "PolTerm platform", identified the term. If the term still proves to be new in the "PolTerm platform" following the manual checking of occurrence thereof, it is analyzed as regards

the syntactic structures in which it usually occurs and the most frequent construction possibilities. If it is not a syntactic structure, or a phrase, but a multiword term that can be extracted from such a construction – it is entered into a preliminary term list. Afterwards the term is subject to conceptual analysis in order to identify its best possible English-language equivalent.

7 Concept analysis

The basic element of terminology work is the term (not concept) as a verbal designation of an appropriate subject-field-related concept.

The process of concept analysis is closely related to term extraction and represents the cognitive dimension of terminology. The information extracted from the textual sources needs to be analyzed from the point of view of domain knowledge structure, which is represented by related concept system expressed by terms.

The cognitive process develops from an object through its generalization and essentialization in our minds to the comprehension of the surrounding reality. It is the process in our consciousness from an image through the meaning represented by the word and through the concept represented by a term to the concept and term systems. See the diagram where the arrows show directions of the processes:

FROM OBJECT TO TERM

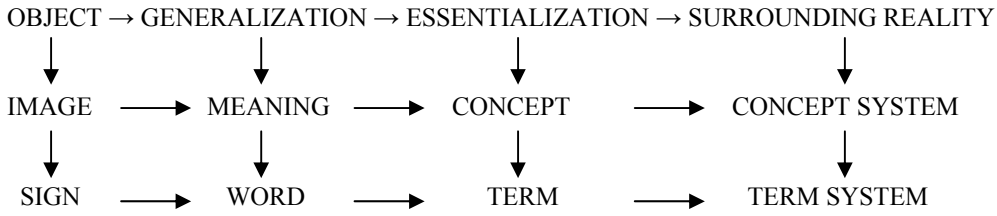

Figure 7.1. The cognitive process from object to term

Conceptual analysis of terms of the appropriate subject field provides a knowledge basis for organizing the concepts into a concept system of the field in question.

Concept analysis has to be based on ISO/TC 37 standards. A concept-oriented approach is applied to the terminology work (instead of word-oriented or context-oriented approach) to ensure that the degree of terminological quality of the work is as high as possible. The concept-oriented approach is relevant when matching terms of different languages for the consolidated EuroTermBank.

Conceptual analysis of the extracted terms is generally carried out in two basic working situations:

- The systematic compilation of the subject field terminology supplemented by conceptual analysis of terms, term groups and resulted in developed subject field term systems represented in term dictionaries, bigger or smaller databases, etc.
- Performance of everyday tasks arising from urgent needs for new terms requested by translators, technical writers, subject field experts, different companies, etc., recently – mainly in connection with the voluminous translations of EU legislation acts, international standards, etc.

Taking into account that the development of science and economy is a continuous process during which new concepts appear and need to be named with new terms, both above-mentioned working situations are daily interconnected. On the one hand, everyday terminology tasks - arising from urgent needs for new terms to express new concepts - lay the foundation for the enrichment of term and concept system of the respective field, and on the other hand, the elaborated term and concept system (of the field in question) alleviates to find a new systemic term for a new concept.

The development of science and economy requires a unified comprehension of actual concepts.

Creation of a concept system includes the following stages, which are more or less actual at all levels (organizational, national, international):

- Determination of the subject field boundaries
- Definition of the mutual relations among concepts
- Classification of concepts (from general to particular ones)
- Testing of the concept content in comparison with other concepts defined in different term vocabularies
- Analyzes of terms (in the linguistic aspect) in different sources
- Preliminary assessment of the developed concept system (the assessment of the term and concept system in the aspect of synonymy, homonymy and polysemy)
- Final assessment of the developed concept system.

The difference is mainly in the scope (size) of the part of the term and concept system in question:

- For only one term and concept
- For a group of related new terms and concepts
- For a whole subject field or subject sub-field in question.

Concept analysis applied in ETB project partner terminology work

ETB partner	**Is/is not applied**
Estonia	In rare cases
Latvia	Usually is applied
Lithuania	Is applied
Hungary	Absence of information
Poland	Consists in cross-system comparative analysis of source legal terms and possible equivalents with the aim of secondary term formation

7.1 Organizational level

Organizations differ by the scope of terminology work:

- From individual terminology tasks to regular terminology work
- From single cases to everyday full-time terminology work
- Without concept analysis at all to concept analysis in the scope of appropriate domain
- Without any staff unit to one or several staff units for terminology work
- From variable staff units to permanent staff group on terminology issues
- From terminology work in the framework of single national language to bi- or multilingual tasks in connection with different translation, etc.

In organizations dealing with terminology issues the contiguity of concept analysis and the ways of solving terminology issues are different, although terminology tasks in many cases are similar or the same.

Existing conditions

7.1.1 Limited framework of a subject field (domain)

- Limited necessity (or lack of necessity) of domain concept analysis
- Without regulation by outside terminology institutions
- Irregular use of term dictionaries (other manuals) and term databases
- Less collaboration with domain specialists and language (or terminology) experts
- Less harmonization of terms used and appropriate concepts
- Possible lack of access to terminology tools

7.1.2 Prerequisites for the best practice

- Usage of term dictionaries and term databases approved by authorized institution, and ask for advice from language and terminology experts
- Involvement of qualified (skilled) translators, linguists and domain experts (for intense terminology work)
- Creation and maintaining at least bilingual term database for internal use
- Analysis of at least a group of related terms and concepts
- Detailed concept analysis for solving problems with closely related concepts that are expressed by homonyms, synonyms or polysemantic terms (at organization level such sets of concepts should be cleared up only in cases when these are directly related to the terms in the documents they are currently dealing with).

If it is necessary to create a new term and if a term is necessary only for local use inside the organization, the main requirement is that a term meets language rules and structural-semantic models of terms.

Dealing with translations, the source language texts are the main base for the compilation of terms, and, if possible, the best experts both of the target language and the appropriate subject field should be involved in solving terminology issues.

In regular translation work, the conceptual analysis is usually done on two levels:

1) On source language level (limited by particular source text to be translated)

2) On target language level for searching the term equivalent via concept analysis and stating the essential characteristics necessary for the concept in question.

7.2 National level

National level of terminology work in connection with the EuroTermBank project means the official level of approved coordinated (unified) terminology in the territory of the particular EU member state as a national term resource for multilingual term database.

Best practice for multilingual term databases that increases the degree of the equivalence of different language terms expressing the same concept is the cross-system comparative analysis of concepts and their designations (terms), or definition-based comparisons, being initially done at national level. It implies dealing with multi-domain and multi-lingual term resources inside each member state and in mutual collaboration of different subject field experts together with high-level linguists and terminologists, if available.

The model of best practice in these activities could be:

- The compulsory status of the official terminology stated in national legislation, e. g. the Official Language Law
- The appropriate governmental regulations providing the compulsory use of the unified and scientifically grounded terminology approved by authorized institution (body)
- An institution (body, e. g. terminology commission) authorized by member state government and founded for concept analysis, decision making, term approval, creating and maintaining national term database, preparation of manuals, instructions, etc. with a status of normative documents.

Some of the relevant principles (see also Chapter *7 Concept analysis*).

- Terms for the national term database are chosen or created on the base of concept analysis of compiled terms, and the concept analysis results in harmonized national multi-branched term system, which provides the high-quality resources for national multilingual term database.
- Concept analysis is necessary to reveal:
    - The types of relationship that hold between concepts, first of all generic, partitive and associative ones
    - Different types of synonyms (including abbreviations, still valid, or in opposite not valid (not preferred) variants, etc.)
    - Term equivalents across two, three or more languages, etc.
- Unification of terms and concepts in the frame of a national language and have to be done before the unification of terms and concepts on the international scale.
- Term synonyms burden precise comprehension of concepts.

Prerequisites

- Involvement of a wide range of subject field experts and broad domain coverage

- Involvement of sufficient number of high level linguists and skilled terminologists
- Close collaboration among these above-mentioned groups
- Implementation of scientific-based requirements for terms and term and concept systems including harmonization on three basic levels – intradomain, interdomain and national
- Stability of skilled (experienced) expert staff and hereditary terminology work
- Unified, regularly updated national term database for outside users in the framework of national member state
- Close collaboration with terminology research institute
- Compulsory terminology course in the professional study programs
- Unified national term database for outside users in a national framework
- Adequate financing

Characteristic feature of the contemporary situation is active communication between different countries and languages. The requirements applied to terms at a national level are actual at international level, too.

Besides, when developing a uniform system of subject field terms and concepts at national level during the translation process a number of specific guidelines have to be taken into consideration, e. g.:

- A term in the source language shall correspond to a single term in the target language for expressing the same concept
- Different terms in the source language expressing different concepts require correspondingly different terms also in the target language
- Such equivalent has to be chosen that in case of back translation the same original term would be used
- The term already established in practice should not be changed without sufficient motivation
- The national term has the preference to international term
- Different attitude to terms which are widely and regularly used in practice: they have to be short, precise, euphonious and easy perceptible; requirements can be mitigated for terms of a more rare usage.

7.3 International level

International level primarily means the observance of appropriate principles accepted for international co-ordination of terminology work. International term and definition standards and other normative acts have to be taken into account at this level. Unfortunately, unconformity of definitions in two-or-more-language term and definition standards used as a basis for international term database resources creates serious discrepancies in the comprehension of one and the same concept.

The relevant principles

- Unification of terms and concepts must be started from the classification of concepts for identification of the main concept groups, subgroups, etc.

- The main attention shall be paid to the unification of the concept level of terms, but the form level depends on the peculiarities of each national language. Any attempt to unify term-forms in all languages would mean unnatural pressure on the national language systems and should be viewed as a destructive and groundless demand. If words in different languages have the same meaning and only different national form ("diverse in form, identical in meaning"), such terms are considered as positive.
- The term should not be translated from one language into another, the equivalent term of a target language must chosen or created (through the concept) to express the same concept of the source language trying to include the same characteristics of the concept in a chosen term.
- One basic language and its terminology must to be chosen as a basis for term and concept analyzes (for EuroTermBank project).
- In cases where the term contradicts its definition (which reflects the concept), the priority should be given to the definition when the term equivalent in the target language is chosen or created.
- It is recommended to take into account the back-translation possibilities.

The concept analysis is necessary for conceptual harmonization of different language terms. It plays a significant role for international term databases. In the framework of EU it is necessary to respect the common concept classification, which may be different in other countries.

Prerequisites (see also national level)

- High quality of term and concept systems (resources?) in each related partner language
- Unambiguous definitions.

8 Tools

The only term extraction tool currently known in ETB partner countries is MorphoLogic's Kilgray. The tool has not yet been released to market and is used in one publishing house for evaluation purposes. According to the description it is a state-of-the-art tool with six algorithms (both rule-based and statistical). Engine to be developed only works for English, German, French, Czech, Polish, and Hungarian at the moment.

9 Data structure and data categories

Irrespective of type of organization, purpose of terminology and type of domain, it is as a principal rule recommended that the data structure permits a broad selection of data categories that enables users to develop entries of a high quality. This does not mean that organizations at different levels should have identical or even similar data structures for storage of terminology work, but within a particular sphere of application the data structure should present a rather exhaustive list of information types. The data structure should furthermore always comply with ISO standards 12200 and 12620 as this will ensure exchangeability and facilitate recognition and comprehension of data categories for new users or outside users.

9.1 Organizational level

Quality is always an important criterion in connection with terminology work, but in an organizational framework speed often plays an important role as well. When these two criteria coexist it speaks in favour of a customized and in some respects more modest data structure with few of the most resource demanding data categories. Another parameter that speaks in favour of a moderately exhaustive data structure is that terminology production in an organizational framework usually is restricted to a very limited number of domains. Consequently, it is easier to select a range of data categories consistent with the requirements of the particular application area.

9.1.1 Data structure

It will usually be recommendable to develop a data structure of 2 to 4 levels dependent on the number of languages involved. If the term collection is monolingual, it is recommended that the data structure contains 2 levels; one level for conceptual information and one level for term related information. Examples of conceptual information are *domain, definition* and *explanation* and examples of term related information are *term* and *context.* This data structure will allow many terms to designate one concept (one definition).

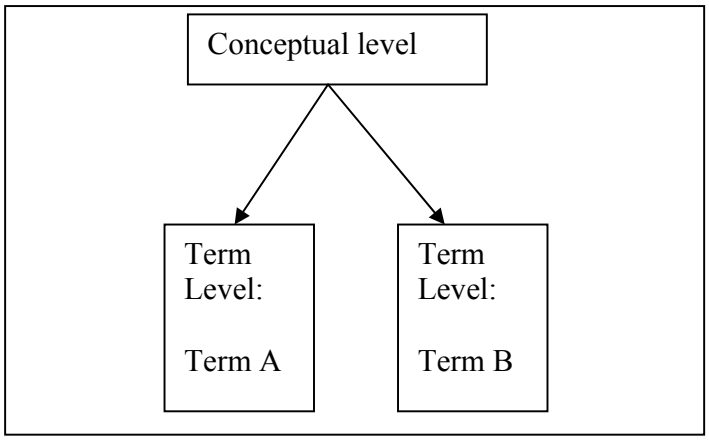

Figure 9.1. Entry structure – monolingual

In connection with a bilingual term collection some options are to let the data structure comprise 2 levels as mentioned above or 3 levels that apart from conceptual and term related information also permit lexical information of the individual words that constitute a term. Whether a *word level* should be added to the data structure depends mostly on the nature of the foreign language.

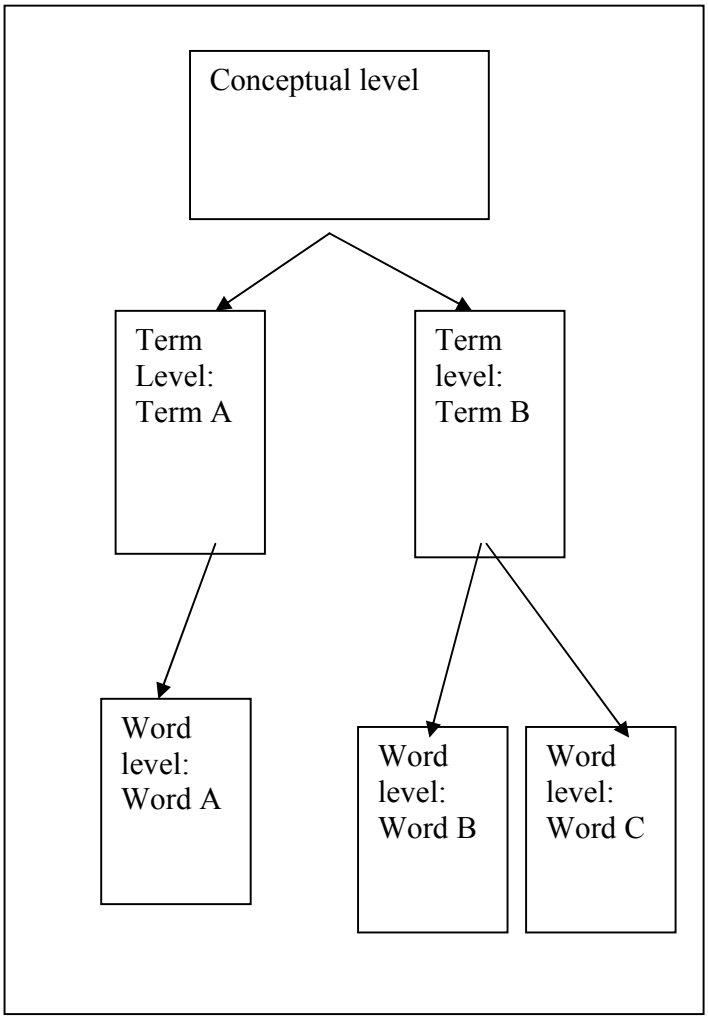

Figure 9.2. Entry structure – bilingual (definition in one language)

A consequence of the data structures described above is however that a definition in only one language can be created for each concept. This may constitute a problem in a bilingual term collection as users may not speak the same native language (and therefore may not fully understand the definition) and as minor conceptual differences must sometimes be expressed. If definitions in both languages are requested, it is recommended to split the conceptual level in two: One level for the language specific information (*language level)* permitting for example a definition for each language and one level for the language independent information *(entry level)* containing for example

domain information (for description of a data structure with 4 levels – which is also recommendable in a multilingual context – see. chapter *9.3.1 Data structure*).

9.1.2 Data categories

It is difficult to recommend a general selection of data categories for the organizational level. Organizations in this level have very different purposes and usually develop terminology in only one or few domains and the data structure should reflect the exact requirements of the particular domains and purposes. There are basic differences between a term collection for technical writers and a term collection for educational purposes regarding linguistic as well as conceptual information.

The below table shows data categories present in all the data sets described in the D.1.1 Report.

Data categories present in all term collections:

Data category	**Information type**	**Description**	**Comments**
Term related	Term	Language 1	
Language related	Definition	Language 1	
Language related	Context/example	Language 1	
Language independent	Subject field	Implicit/explicit	When a term collection covers only one domain, this info is implied
Administrative	Administrative	Serial number/ entry identifier/ author code, etc.	Various subtypes are used

9.2 National level

In a national framework it is recommended that the data structure permits an exhaustive selection of information types that cover very different user requirements and enable users to develop entries for very different purposes and of a very high quality. It is in other words a reasonable strategy to opt for a maximum solution covering all relevant data categories for all conceivable situations of the particular terminology work.

It is also recommended to let pragmatics contribute to the picture and only establish a limited selection of obligatory data categories. Too many obligatory data categories may become a curb on terminology production and too few will influence the quality level. It is recommended that the obligatory data categories include some administrative information as date of entry creation and author information and in addition to this at least subject information (for example classification number and classification system), the term and a reference to the term.

It is essential that the data structure is based on standards to ensure exchangeability with other data collections and to ensure that data categories are recognizable for outside users. It is recommended that the data structure comply with ISO standards 12200 and 12620 (this is recommended for all levels).

A condition which also plays an important role in selection of data structure is that term bases at national level usually are monolingual or bilingual.

9.2.1 Data structure

The data structure recommended for a monolingual term base at the national level is the same as the one recommended for the organizational level, i.e. the data structure should comprise 2 levels covering conceptual and term related information (cf. chapter *9.1.1 Data structure*). Similarly, the data structure in a bilingual term base should comprise at least conceptual and term related information and possibly lexical information as well. Furthermore, it should be considered whether it is necessary to split the conceptual level into two levels in order to permit a definition for each language (cf. chapter *9.1.1 Data structure*).

9.2.2 Data categories

For an inventory of possible data categories see descriptions for international level chapter *9.3.2 Data categories.*.

9.3 International level

At the international level the criteria considered of primary importance are the same as for the national level. The data structure should present an exhaustive selection of data categories in order to permit high-quality entries and ISO standards should be observed in order to ensure exchangeability as this is an essential factor.

Another factor demanding special attention in an international terminology cooperation is user requirements. Though it is always important to base a data structure on user requirements, it is especially essential to involve user representatives actively in an international framework as requirements will necessarily be differentiated from one national organization to another.

The nature of an international terminology cooperation implies that term bases at this level usually are multilingual.

The below data structure recommended for the international level is partly inspired by the IATE database.

9.3.1 Data structure

It is recommended that the data structure comprises information about the concept, the terms that designate the concept and the words that constitute the individual terms. As a term base in an international cooperation will usually be multilingual, it is recommended that the data structure permits definitions in all languages and therefore conceptual information should be grouped in two levels: the *entry level* containing language

independent information and the *language level* containing language specific information. Term related information should be contained at *term level*; an example of an information type that might appear at term level is usage information. Lexical information concerning a specific word should be contained at *word level*.

The below figure illustrates the recommended overall structure of an entry.

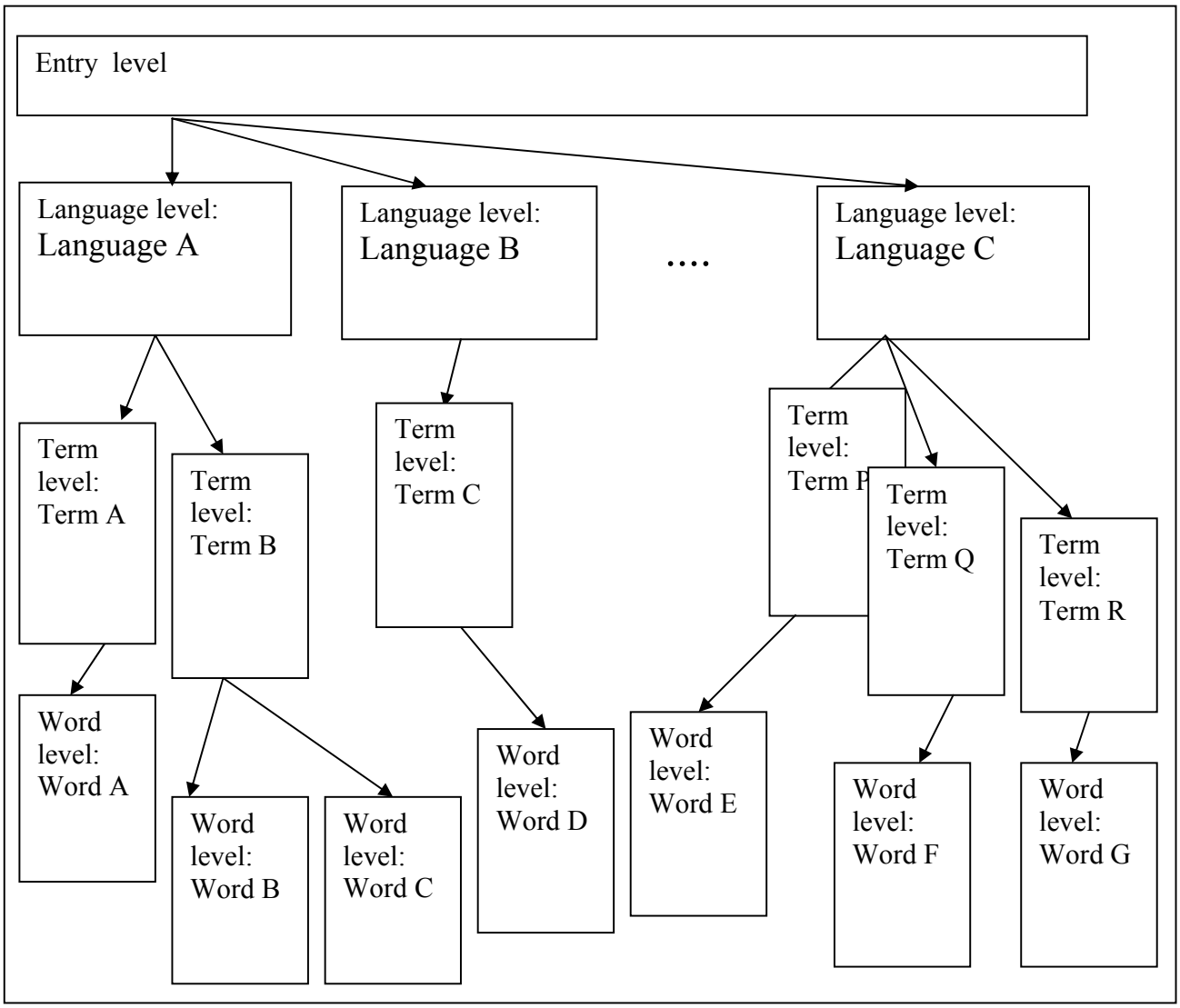

Figure 9.3. Entry structure - multilingual

9.3.2 Data categories

Below are described the data categories which are suggested for each level of an entry in an international framework. The organization of data categories is by level, i.e. if a data category can appear at several levels, it is repeated for each of these levels. For some data categories its position code in ISO 12620 is mentioned, in connection with other data categories a specific position code is not mentioned as ISO 12620 contains multiple

possibilities that must be considered in relation to the specific application. Note that this is by no means an exhaustive list of data categories contained in ISO 12620. Please consult the ISO standard for inspiration and more details.

Data categories suggested as obligatory are the same as for the national level (cf. chapter *9.2 National level*).

9.3.3 Entry level

Administrative information

Entry identifier; 12620 Position Code: A.10.15
The value of this data category is a system-generated number that will identify the entry uniquely.

Subset owner; 12620 Position Code: A.10.02.02.10
The value of this data category is the institution responsible for the whole entry. As the data collection in an international framework will contain contributions from many different organizations it is necessary to state clearly who is responsible of maintenance of each entry.

Originator; 12620 Position Code: A.10.02.02.01
An identifier of the person who prepared the entry.

Inputter; 12620 Position Code: A.10.02.02.02
An identifier of the person who types in the information.

Origination date; 12620 Position Code: A.10.02.01.01
The date the entry was first created.

Updater; 12620 Position Code: A.10.02.02.03
The value of this field is the person having made the latest changes to the information at entry level.

Modification date; 12620 Position Code: A.10.02.01.03
The date when the latest changes to the entry level were made.

Security subset; 12620 Position Code: A.10.03.09
This data category contains a security classification expressing the confidentiality level of the entire entry. A security classification can be used in connection with for example critical terms during a development phase

Subject information

Subject information
The data category(ies) chosen for subject information will contain the domain of the particular concept.

Note; 12620 Position Code: A.08

A free descriptor field to allow for other kinds of subject information that cannot be expressed in the subject information field(s).

Non-textual information
Non-textual information
The data category(ies) chosen for non-textual information will contain for example tables, figures, videos and other binary data.

Reference
Reference(s) to the non-textual information.

Collection
Data collection
This field can be used to signify that a particular concept belongs in a particular collection of concepts.

Source language
Source Language
This information concerns the source language of a set of terms that are not perfectly multi-directional. There is currently no 12620 data category to indicate the source language in a set of terms that are not perfectly multi-directional, but there are some alternative possibilities that can be considered.

Cross-reference information
Cross-reference information
A reference to other concepts in various ways related semantically to the concept in question, for example broader concept, subordinate concept or related concept.

9.3.4 Language level

Administrative information
Originator; 12620 Position Code: A.10.02.02.01
An identifier of the person who prepared the language level.

Inputter; 12620 Position Code: A.10.02.02.02
An identifier of the person who types in the information.

Origination date; 12620 Position Code: A.10.02.01.01
The date the language level was first created.

Updater; 12620 Position Code: A.10.02.02.03
An identifier of the person having made the latest changes to the information at language level.

Modification date; 12620 Position Code: A.10.02.01.03
The date when the latest changes to the language level were made.

Language symbol; 12620 Position Code: A.10.07
This data category contains the language symbol of the particular language. The symbols specified in ISO 639 should be used.

Non-textual information
Non-textual information

Reference

See comment about non-textual information at entry level.

Specification of the concept
Definition; 12620 Position Code: A.05.01
In this field, a formal and precise description of the concept is given.

Reference
Reference(s) to where the definition given above was found.

Explanation; 12620 Position Code,: A.05.02
Compared to the *Definition* field, this field makes it possible to give a more informal description of the concept. This field would be particularly useful in cases where a formal definition has not been obtainable.

Reference
Reference(s) to where the explanation given above was found.

Note; 12620 Position Code: A.08
This data category can contain some additional and general information about the concept in the particular language or the field can contain information related to the definition or explanation.

Reliability
Reliability code; 12620 Position Code: A.03.04

Reliability codes are suggested at language and term levels. A reliability code at the language level will thus provide an assessment of the correctness and precision of the information given in relation to the specific concept.

9.3.5 Term level

Administrative information
Originator; 12620 Position Code: A.10.02.02.01
An identifier of the person who prepared the term level.

Inputter; 12620 Position Code: A.10.02.02.02
An identifier of the person who types in the information if this person varies from the originator.

Origination date; 12620 Position Code: A.10.02.01.01
The date the term level was first created.

Updater; 12620 Position Code: A.10.02.02.03
An identifier of the person having made the latest changes to the information at term level.

Modification date; 12620 Position Code: A.10.02.01.03
The date when the latest changes to the term level were made

Entry source; 12620 Position Code: A.10.13
If the entry is imported from another resource this field will always contain information about the database or format from which data are imported.

Search term; 12620 Position Code: A.10.06.03
This field will contain related forms of the term to facilitate searching. The author of term level information containing a verb may e.g. expect that users will often make a search for the adjectival form. In this case the author can state the adjectival form in *search term*

Terms
Term; 12620 Position Code: A.01
This field will contain the term: a designation of a defined concept in a specific language by a linguistic expression.

Term Type; 12620 Position Code: A.02.01
The value in the *Term Type* field is an attribute assigned to a term. The values can be selected from a picklist containing the term types used by the organizations. A picklist for *termtype* is contained in ISO 12620.

Reference
Source(s) of the *term*.

Usage information
Usage information
Data categories selected for usage information may for example concern a textual example of a concrete use of the term in question, a classification indicating the relative level of language of a term, information about the use of a particular term over time, the status of a term with respect to standardization etc.

Note; 12620 Position Code: **A.08**
A general comment that applies to the entire term level.

Reliability
Reliability code; 12620 Position Code**:** A.03.04

Reliability codes are suggested at language and term levels. A reliability code at the term level will thus provide an assessment of the correctness and precision of the information given in relation to the specific term

Validation
Validation information
It is suggested that validation information is located at term level and not at the other levels though a validation procedure includes validation of all levels.

Validation information may for example include identifiers of persons checking and approving entries together with relevant dates. In an international framework it may however be necessary to record a more complex validation procedure with several validation stages. Data categories reflecting a complex validation procedure are not contained in ISO 12620.

9.3.6 Word level

As a term may be a multiword string this level is created to contain information that concerns the individual words of a term.

Administrative information
This level can contain the following administrative fields:

Originator; 12620 Position Code: A.10.02.02.01
An identifier of the person who prepared the word level.

Inputter; 12620 Position Code: A.10.02.02.02
An identifier of the person who types in the information.

Origination date; 12620 Position Code: A.10.02.01.01
The date the entry was first created.

Updater; 12620 Position Code: A.10.02.02.03
An identifier of the person having made the latest changes to the information at word level.

Modification date; 12620 Position Code: A.10.02.01.03
The date when the latest changes to the word level were made.

Word
*Term element***;** 12620 Position Code: A.02.08.02
This data category concerns a particular word that forms part of a term.

Lexical information
Dependent on the languages involved in the international cooperation some data categories for grammar information should be selected. Data categories for lexical

information are for example, part of speech, grammatical number, grammatical gender etc.

Pronunciation
Dependent on involved languages and purpose of terminology, pronunciation information may be necessary.

*Pronunciation***;** 12620 Position Code; **A.02.05**
This data category contains a representation of the pronunciation of a word.

10 Exchange Format

The creation of high-quality terminology is both time-consuming and cost-intensive. As a consequence, the community of terminology users has a vested interest in exchanging terminological data collections. Different user-group needs and organizational environments dictate, however, that the languages and information categories required by individual systems will vary considerably, which means that the structure of different terminology databases will also exhibit a great deal of diversity. This complication applies even in cases where the individual systems are themselves relatively simple. As a result, any exchange of terminological data between different systems becomes significantly more difficult than one might anticipate. In the past, these problems have made it necessary for exchange partners to create individual conversion programs to accommodate each exchange situation.

In order to overcome these costly individual programming of conversion routines, ISO/TC 37 has developed three international standards related to terminology interchange. ISO 12200 and ISO 12620 (see Deliverable 1.1 *Current standards and best practices assessment report*, chapter 9.1.4 and 9.1.5) specify the MARTIF interchange format and the corresponding data categories, but these two standards from 1999 only allow for negotiated interchange and are not strict enough for a specific interchange scenario without additional agreements. ISO 16642 (see Deliverable 1.1, chapter 9.1.6) is related to the terminology markup framework TMF enabling to specify interoperable markup languages on the basis a common meta model. Therefore, TMF is not a terminology interchange format in itself, but MARTIF is such a TMF-compatible markup language.

LISA, the Localization Industry Standards Association, has developed and specified TBX (see Deliverable 1.1, chapter 7.8.1), a very practical terminology exchange format that is compliant with the terminology markup framework TMF. It can be assumed that many developers of terminology management tools and other language processing applications will support TBX as an exchange format in the near future. Therefore TBX must be the recommended exchange format for terminological data in almost every specific interchange scenario.

TBX (TermBase eXchange) is an open XML-based standard format for terminological data. It provides a number of benefits as long as TBX files can be imported into and exported from most software packages that include a terminological database. This capability facilitates the flow of terminological information throughout the information cycle both inside an organization and with outside service providers. In addition, terminology that is made available to the general public should become much more accessible to humans and more easily integrated into existing terminological resources.

For various types of machine processing, including transmission over the Internet, terminological data can be represented using XML. The TBX format is a standard-based XML application designed to support machine processing of terminological data in

various computer environments, including standalone computers, the Internet, and intranets.

The terminological framework for TBX is provided by three established international standards: ISO 12620, ISO 12200, and ISO 16642.

TBX is designed to support the analysis, representation, dissemination, and exchange of information from human-oriented terminological databases (termbases). TBX is a format that qualifies as a TML (Terminology Markup Language) by complying with the requirements of the Terminology Markup Framework (TMF) according to ISO 16642. It is based on the TMF structural metamodel; it specifies a set of data categories from ISO 12620 and adopts an XML style compatible with ISO 12200.

Each variant of TBX is a TML within TMF. Since each TML is interoperable with every other TML, limited only by incompatibilities in the choice of data categories, TBX XML documents can be converted to XML documents in other formats within TMF. However, interoperability between TBX and formats that do not qualify as TMLs is not guaranteed. Nevertheless, limited interoperability is possible between non-TML formats such as OLIF and TMX.

Even though TBX supports customization according to user needs, there are limits to what variations can be defined by an XCS file; otherwise, certain variations would not qualify as TMLs according to 16642. All acceptable variations on TBX have the same core structure. They differ mainly with respect to the data categories from ISO 12620 that are allowed by a particular user group.

According to the hierarchy of a TBX document, the highest-level XML element is the *martif* element, which contains a <*martifHeader*> element and a <*text*> element. The <*martifHeader*> element provides a description of the file, on the applicable XCS file and unusual character encoding, and a history of major revisions to the collection.

The <*text*> element contains the terminological data. It includes in the <*body*> the actual terminological entries – one entry per concept – enclosed in <*termEntry*> tags, as well as complementary information, e. g. bibliographical data, in the <*front*> and <*back*> elements, to which can be referred from the <*body*> entries. Within the terminological concept entries various data categories allow to provide different kinds of information, either in free text or chosen form a pick list, as well as cross-references that points to either somewhere inside the *martif* element or to an external object using a URL. The terminological concept entries *(*<*termEntry*>) can be multi- or monolingual.

```
<?xml version='1.0'?>
<!DOCTYPE martif SYSTEM  "./TBXcoreStructureDTD-v-1-0.DTD">
<martif type='TBX' xml:lang='en' >
<martifHeader>
    <fileDesc>
        <sourceDesc><p>from an Oracle corporation termBase</p>
```

```
        </sourceDesc>
    </fileDesc>
    <encodingDesc><p type='DCSName'>TBXdefaultXCS-v-1-0.XML</p>
    </encodingDesc>
</martifHeader>
<text> <body>
    <termEntry id='eid-Oracle-67'>
        <descrip type='subjectField'>manufacturing</descrip>
        <descrip type='definition'>A value between 0 and 1 used in …</descrip>
        <langSet xml:lang='en'>
            <tig>
                <term tid='tid-Oracle-67-en1'>alpha smoothing factor</term>
                <termNote type='termType'>fullForm</termNote>
            </tig>
        </langSet>
        <langSet xml:lang='hu'>
            <tig>
                <term tid='tid-Oracle-67-hu1'>
                    Alfa sim&#x00ED;t&#x00E1;si t&#x00E9;nyez&#x00F5;
                </term>
            </tig>
        </langSet>
    </termEntry>
</body> </text>
</martif>
```

Figure 10.1. Example of a TBX document

TBX includes meta-markup tags for distinguishing embedded non-TBX markup from text. They allow TBX elements to contain various kinds of other markup, e. g. html or text processing markup that needs to be retained but should not necessarily be processed during terminology management functions.

In the Annex and separate files of the TBX specifications, important information and examples for the encoding of data according to TBX is provided. This includes a formal XML representation of the core structure containing the basic data categories, element groups, attribute lists and comments associated with meta data, tables defining the TBX master XCS (data constraint specification), and guidelines for encoding particular data categories in TBX (e.g., as XML elements).

The TBX format is a principle result of the SALT project (see chapter 6.2.2). The LISA[1] OSCAR[2] group has adopted TBX as its terminology exchange format and continues the development of the specifications.

[1] Localization Industry Standards Association
[2] Open Standards for Container/Content Allowing Re-use

10.1 Organizational level

The exchange of terminological data within a specific organization is very simple, if only one type of terminology management system with a unique entry structure is applied. Data can be exchanged either by using simple formalisms like comma-separated files or the system-specific exchange format. But if different termbases (for specific user groups or applications) with different data categories and entry structures exist within one organization, the exchange problematic is much more complex. A customized specific exchange routine between two termbases can be programmed, but the more terminology resources are involved the more additional exchange routines are necessary.

Although such a customized format can solve all needs for terminology interchange with a specific organization, there is a strong recommendation to also use standardized exchange formats like TBX for this individual exchange scenario, since sooner or later new termbases (or other applications) will come along and the need for terminology interchange with other organizations will arise.

10.2 National level

On national level, only a standardized format for the exchange of terminological data can be recommended, because all systems and partners involved in the exchange process have to refer to a well defined, widely known, and appropriate format specification like TBX.

10.3 International level

The recommendations for terminology interchange on international level are not very different from the situation on national level. The multilingual aspect of terminology resources to be exchanged requires no additional demands to the exchange format in comparison to monolingual or bilingual resources. Only the fact that more partners with different terminology management systems and different term base structures may be involved in the exchange process gives a stronger impact to the requirement to use a standardized exchange format like TBX.

On the basis of the detailed analysis of exchange formats described in Deliverable 1.1 and the arguments described in 7.8.1, we strongly recommend to use **TBX (TermBase eXchange)** as the exchange format for terminological data on the organizational, national and international level.

11 Management of terminological entries

11.1 Organizational level

At the organizational coordination level, management of terminological entries may often meet process requirements less demanding than those described for the national level (cf. chapter *11.2 National level*). In this comparison also the scope of responsibilities for particular persons and bodies as well as search and retrieval options are different.

11.1.1 Existing conditions of term collections

The presently existing conditions of terminology work at the organizational level show that not all relevant partners have access to internet and indispensable terminology tools, not always involve language experts, terminologists and domain experts in addition to translators. They do not cover a broad scope of domains, nor cooperate with relevant institutions in terminology coordination, because their (usually bilingual) databases are seldom multilingual, and their financial situation is not always satisfactory to such an extent that they are able to use better terminology tools or employ more experts.

11.1.2 Basic differences

Basic differences between the organizational level and national level show that terminology work performed at the national level is usually much more demanding in many aspects. The requirements that not always must be met at the organizational level include the permanent cooperation with terminologists, domain experts and language experts, cover more than one domain and consult national institutions, which is not the case at the organizational level.

11.1.3 High priority criteria

There are the following criteria against which management of terminological entries at the organizational level can be assessed:

1. Easy use and maintenance of a database
2. Speed of the database updating
3. Bilingual nature of a database
4. Availability to in-house users
5. Quality of terminological entries according to the specific needs
6. Harmonization of terminology within a collection
7. Availability to external users
8. Availability of an option to exchange terminology between terminology collections.

11.1.4 Process requirements

Process requirements specify what are the priorities of particular partners in striving for effective management of their databases at the organizational level. The partners have agreed on easy use and maintenance of a database (1), the necessary speed of the database updating and bilingual nature of their databases (3) that do not need to be multilingual at this level.

It is also important to have the terminology database available to in-house users (4) such as translators as well as domain experts and terminologists, if any. The quality of terminological entries, however, do not need to comply with the theoretical principles, if there are other specific preconditions, such as the compatibility with other important databases (e.g. Eurodicautom).

Managers of terminology databases at the organizational level are not always able to provide for harmonization of terminology within a collection (6), availability to external users (7) or availability of an option to exchange terminology between terminology collections (8) as distinguished from the national and international levels.

11.1.5 Responsibilities of experts

In addition to translators and a terminology coordinating translator, other employees may be not involved in management tasks at the organizational level.
Although the involvement of relevant professionals is an indispensable guarantee of the high quality of terminological entries covered by a term database and their efficient management, but at this level they are often absent.

Nevertheless, if such specialists are involved, their responsibilities are assigned to:
1. A terminologist
2. A domain expert or experts
3. A language expert.

At the organizational level, a terminologist, permanently provided with consultation of domain experts, is responsible for making relevant decisions. Management activities on this level are also carried out by the terminologist and domain experts and a language expert, if involved, which is not always the case at this level.

11.1.6 Search and retrieval options

The prerequisite of these management tasks is that terms are extracted by a terminologist from machine-readable texts and, adequately formatted according to the specific structure of data categories, are automatically added (imported) to a database in the compilation process as new entries. There is no difference in this respect in comparison with the national level.

The following are the search and retrieval options indispensable for term databases at the organizational coordination level:

1. Basic retrievability options enabling to search a term database, as seen from database owners' perspective, on the basis of any pre-defined criterion or a set of criteria as above
2. Basic retrievability options enabling to search a term database, as seen from external users' on the basis of any pre-defined criterion or a set of criteria attributable to recorded terminological information entered in all data categories

of entries containing such terms, such as: a subject field, a source, a string of letters contained in a term, definition words, or an origination date
3. Protection devices, once the options are limited to users, which may be not allowed, for example, to export a set of terms retrieved from a database according to a criterion, to modify, update or delete an entry
4. Convertibility of database data into other formats, so that the exchange of data between different collections be easily done.

All the partners are able to use the basic retrievability options mentioned under 1, 2 ,3 and 4.
These basic requirements have to be met also at the organizational level (cf. chapter *11.1.4 Process requirements*) and differences have not been found in this comparison.

11.2 National level

On the national coordination level, management of terminological entries must meet much higher process requirements than those described for the organizational level (cf. *11.1.4 Process requirements*). In this comparison also the scope of responsibilities for particular persons and bodies as well as search and retrieval options are different.

11.2.1 Existing conditions of term collections at the national level

The presently existing conditions of terminology work at the national level show that all relevant partners have access to internet and indispensable terminology tools, involve not only translators, but also language experts, terminologists and domain experts, cover a broad scope of domains, cooperate with relevant institutions in terminology coordination in both mono- and bilingual term collections, and their financial situation is satisfactory to such an extent that the regular terminology work is continued.

Main differences between the organizational level and national level show that terminology work performed at the national level is usually much more demanding in many aspects. The requirements that not always must be met at the organizational level include the permanent cooperation with terminologists, domain experts and language experts, cover more than one domain and consult national institutions, which is not the case at the organizational level.

The below considerations are relevant to the following ETB partners:

- Estonia: The Estonian Legal Language Centre
- Latvia: The Terminology Commission of the Latvian Academy of Sciences
- Lithuania: The State Commission of the Lithuanian Language
- Hungary: The (Hungarian) Ministry of Justice
- Poland: The (Polish) Office of the European Integration Committee.

11.2.2 High priority criteria

There are the following criteria against which management of terminological entries can be assessed:

1. High quality of terminological entries covered by a term database created according to the research principles and their efficient management
2. Exchangeability between terminology collections
3. Availability to external users
4. Comprehensive and multifunctional design
5. Easy use and maintenance of a database
6. Harmonization of terminology within a collection
7. Monolingual or bilingual nature of a database.

11.2.3 Process requirements

Process requirements specify what are the priorities of particular partners in striving for effective management of their databases at the national level. All the partners have agreed on the high quality (1), availability to external users (3), easy use and maintenance of a database (5), harmonization of terminology within a collection (6) and at least bilingual nature of their databases.

Not all of them are able to provide for either the option of exchangeability between other terminology collections (2) or a comprehensive and multifunctional design (4), which is also a case of term databases at the organizational level (cf. chapter *11.1.4 Process requirements*).

The below table shows the attitude of the partners towards the above mentioned priority criteria.

ETB partner	Process requirements criteria
Estonia	1, 2, 3, 5, 6, 7;
Latvia	1, 3, 4, 5, 6, 7;
Lithuania	1, 3, 4, 5, 6, 7;
Hungary	1, 2, 3, 5, 6, 7;
Poland	1, 3, 5, 6, 7.

Therefore the best practice provides solutions to all these requirements but exchangeability and comprehensive/multifunctional design.

11.2.4 Responsibilities of experts

Responsibilities define which types of employees are involved in the different types of management tasks in order to ensure a proper management of work and information flow.

The involvement of relevant professionals is an indispensable guarantee of the high quality of terminological entries covered by a term database and their efficient management.

Such responsibilities are assigned to:

1. A terminologist or a group of terminologists with a database development (i.e. technical) and subject-field(s) expertise

2. A domain expert or a number of experts
3. A translator or a group of translators
4. A language expert or a group of experts.

At the national level, a terminologist or a group of terminologists, permanently provided with consultation of domain experts, are responsible for making relevant decisions. Management activities on this level are also carried out by the same persons: a terminologist or a group of terminologists cooperating with domain experts from relevant national institutions mentioned above, translators and language experts, which is not always the case at the organizational level (cf. chapter *11.1.5 Responsibilities of experts*).

The below table shows the best practice as regards particular responsibilities:

ETB partner	Responsibilities
Estonia	1, 2, 3, 4 ;
Latvia	1, 2, 3, 4;
Lithuania	1, 2, 3, 4;
Hungary	1, 2, 3, 4;
Poland	1, 2, 3, 4.

11.2.5 Search and retrieval options

Search and retrieval options define how the complexity of management tasks carried out by those responsible for database contents is reflected in the complexity and number of search and retrieval options for external users. As these operations are intertwined, it is necessary to report on them from two perspectives: that of terminologists responsible for the database content and that of external users of the database.

The prerequisite of these management tasks is that terms are extracted by terminologists from machine-readable texts and, adequately formatted according to the specific structure of data categories which should differ according to scenarios (cf. chapter *9 Data structure and data categories*), are automatically added (imported) to a database in the compilation process as new entries. In some databases it is also users that can import new entries.

The following are the search and retrieval options indispensable for term databases at the national coordination level:

1. Basic retrievability options enabling to search a term database, as seen from database owners' perspective, on the basis of any pre-defined criterion or a set of criteria as above.
2. Basic retrievability options enabling to search a term database, as seen from external users' on the basis of any pre-defined criterion or a set of criteria attributable to recorded terminological information entered in all data categories of entries containing such terms, such as: a subject field, a source, a string of letters contained in a term, definition words, or an origination date.

3. Protection devices, once the options are limited to users, which may be not allowed, for example, to export a set of terms retrieved from a database according to a criterion, to modify, update or delete an entry.
4. Convertibility of database data into other formats, so that the exchange of data between different collections be easily done.

All the partners are able to use the basic retrievability options mentioned under 1, 2 ,3 and 4.
These basic requirements have to be also met at the organizational level and no difference have been found in this comparison.

The below table shows the above described options available to the particular partners:

ETB partner	Search and retrieval
Estonia	1, 2, 3, 4;
Latvia	1, 2, 3, 4;
Lithuania	1, 2, 3, 4;
Hungary	1, 2, 3, 4;
Poland	1, 2, 3, 4.

11.3 International level

On the international coordination level, management of terminological entries must meet higher process requirements than those described for the organizational level, but not much different than those provided for the national level. In comparison with the latter the scope of responsibilities for particular persons and bodies as well as search and retrieval options are seldom different.

11.3.1 Existing conditions of term collections

The presently existing conditions of terminology work at the international level show that all relevant partners have access to internet and indispensable terminology tools, involve not only translators, but also language experts, terminologists and domain experts, cover a broad scope of domains, cooperate with relevant institutions in terminology coordination in both mono- and bilingual term collections, and their financial situation is satisfactory to such an extent that the regular terminology work is continued.

Not many differences exist between the national level and international level: at the both levels terminology work performed at the international level is also demanding in many aspects. The requirements include the permanent cooperation with terminologists, domain experts and language experts, their term databases cover many domains and require consultations with national institutions in different countries, sometimes with international organizations.

The below considerations are relevant to two exemplary organizations:

International Standardization Organization ISO, and
Inter-institutional Term Databases for EU Institutions and Agencies IATE.

11.3.2 High priority criteria

The criteria against which management of terminological entries can be assessed at this level are different only in some respects as compared with the criteria at the national level. The below list is identical in this comparison in items 1 to 6, but due to the multilingual and multi-domain nature of term databases at the international level, it is necessary to provide for inter-language and inter domain coordination (items 7 and 8).

1. High quality of terminological entries covered by a term database created according to the research principles and their efficient management
2. Exchangeability, meaning the availability of an option to exchange terminology between terminology collections
3. Availability to external users
4. Comprehensive and multifunctional design
5. Easy use and maintenance of a database
6. Harmonization of terminology within a collection
7. Multilingual nature of a database
8. Multi-domain content of a database.

11.3.3 Process requirements

Process requirements specify the priorities of terminology managers to obtain efficiency at the international level. Term databases at this level must be of high quality (1), exchangeable with other databases (2), available to external users (3), have a comprehensive and multifunctional design (4), must be easy in use and maintenance of their database (5), provide for harmonization of terminology within the database (6) be multilingual (7) and multi-domain (8).

These requirements are still more demanding than those envisaged for the national level where their databases are seldom multilingual and multi-domain in nature.

11.3.4 Responsibilities of experts

Responsibilities of employees involved in the different types of terminology management tasks are almost the same as those described at the national level (cf. chapter *11.2.4 Responsibilities of experts*).

The involvement of relevant professionals is an indispensable guarantee of the high quality of terminological entries covered by a term database and their efficient management.

Such responsibilities are assigned to:
1. A terminologist or a group of terminologists with a database development (i.e. technical) and subject-field(s) expertise
2. Domain experts and inter-domain coordination experts
3. Translators

4. Language experts and inter-language coordination experts.

At the international level, terminologists, permanently provided with consultation of domain experts, are responsible for making relevant decisions. Management activities on this level are also carried out by the same persons: terminologists cooperating with domain experts from relevant national institutions in different countries, translators and language experts.

11.3.5 Search and retrieval options

Search and retrieval options define how the complexity of management tasks carried out by those responsible for database contents is reflected in the complexity and number of search and retrieval options for external users. The options are quite similar to those which must be ensured at the national level.

The following are the search and retrieval options indispensable for term databases at the international coordination level:

1. Basic retrievability options enabling to search a term database, as seen from database owners' perspective, on the basis of any pre-defined criterion or a set of criteria as above
2. Basic retrievability options enabling to search a term database, as seen from external users' on the basis of any pre-defined criterion or a set of criteria attributable to recorded terminological information entered in all data categories of entries containing such terms, such as: a subject field, a source, a string of letters contained in a term, definition words, or an origination date
3. Protection devices, once the options are limited to users, which may be not allowed, for example, to export a set of terms retrieved from a database according to a criterion, to modify, update or delete an entry
4. Convertibility of database data into other formats, so that the exchange of data between different collections be easily done.

The partners at this level are able to use the basic retrievability options mentioned under 1, 2, 3 and 4. Since these requirements are met by them in practice, they must be recommended as appropriate for the international level.

12 Validation workflow

The validation of newly created terminology is essential in order to guarantee a satisfactory quality. What "satisfactory" would be, has to be defined by every terminology creating organization according to its specific process requirements. Whereas high quality seems to be obviously the most important criteria, financial and time constraints can force an organization to cut back quality requirements.

The validation of term entries consists of two main steps concerning the formal structure on the one hand and the content on the other hand. Two criteria are decisive for the complexity of the validation workflow and, thus, for the time and budget spent for it: the mono- or multilingual orientation of terminology work and the amount and qualification of the people involved. The verification of terminology in many languages carried out by experts from different countries requires a huge coordination effort.

The different scenarios for the description of best practice for validation workflow are based on the distinction between international, national, and organizational levels. Whereas validation processes are similar on international and national levels, the conditions and aims of terminology creation are much different on organizational level, and so is the validation workflow. However, if the conditions assumed for national/international level are applicable to an organization, the characteristic steps, actors and technical means of the validation procedure are transferable to the organizational level, and vice versa.

12.1 Organizational level

12.1.1 Process requirements

On organizational level a restricted budget and a tight timeframe is more likely than on national or international level. The ranking and requirements for terminology work can all the same differ widely as they are rather linked to the nature of the organization's core business. So, terminology can be an important issue for an organization or a "necessary evil", it can be mono-, bi- or multilingual, used for internal processes – stock-keeping and the like – or for customer relations, for standardisation purposes, for a special customer or project, etc. Big companies may assign more financial support to terminology, even if it is not part of the core business, than SME.

For the purpose of a schematic representation we assume that, on organizational level, terminology work is mono- or bilingual and covers a limited number of subject fields. It is performed by a small number of in-house employees without special terminology management tools. Beside quality, speed of the terminology work and up-to-dateness of the data are just as important.

The assessment of the formal correctness of the terminology entries contains the following checks:

- Check for duplicates

- Integrity check – completion of all mandatory fields, no double completion of fields to be completed only once
- Format control – e. g. date formats, references, etc.
- Consistency check – correct form of the terms, e. g. singular for substantives, infinitive for verbs, case sensitivity, ISO language codes
- Spell check
- Grammatical check – correctness of the grammatical information to every term.

In case of pure word lists or databases defined in a rather simple way, the validation of the formal structure is even easier consisting, for example, only of the check of a possible lack of terms in every language covered. But even without terminology management tools available, terminology developers shall meet some minimal formal requirements, also with regard to the possible need of an exchange routine.

The second, and far more complicated, step contains the content check for every entry. It shall comprise the verification of

- Choice of terms in every language according to predefined criteria, like linguistic correctness (i. e. ISO 704)
- Correctness of synonyms
- Exactness of definitions as regards content
- Correctness of graphical representations
- Correctness of usage notes, temporal qualifiers, register, subject field, etc.,

always provided that the database contains the respective information. In most cases not all data categories might be covered.

The formal correctness of an entry should be a minimum requirement for a terminological database and can be accomplished in a simple way. But also the content check shall be mandatory to organizations of all levels, even if the data is for internal use only.

12.1.2 Actors

On organizational level the validation routine is in general carried out by internal employees. The maintenance of the terminology database shall be assigned to a translator, if no terminologist is involved. In an ideal case the creator of an entry must not control his own work, although this might be inevitable due to a lack of qualified staff. The formal check shall be performed by a translator without subject field-specific knowledge, if possible assisted by technical means like an automatic spell check.

The content validation shall be carried out by a subject field expert; in most cases this would be an internal expert working in the respective department of the organization

The person responsible for the maintenance of the terminology database shall develop an easy marking system appropriate for identifying the state of the entries containing at least two tags to identify entries already validated and those not yet finalised.

12.1.3 Feedback mechanisms

Validation processes carried out by in-house staff allow for rather informal and quick response procedures. The feedback can be provided even orally, per e-mail or in paper form; depending on the amount of terminology created and on the size of the organization, a predefined feedback form would be helpful.

The translator maintaining the database modifies the entries according to the experts' comments and, if necessary, after direct consultation with them. Once the entries tagged as valid they shall be made available to the colleagues concerned. This might happen when and as the need arises or on a regular basis.

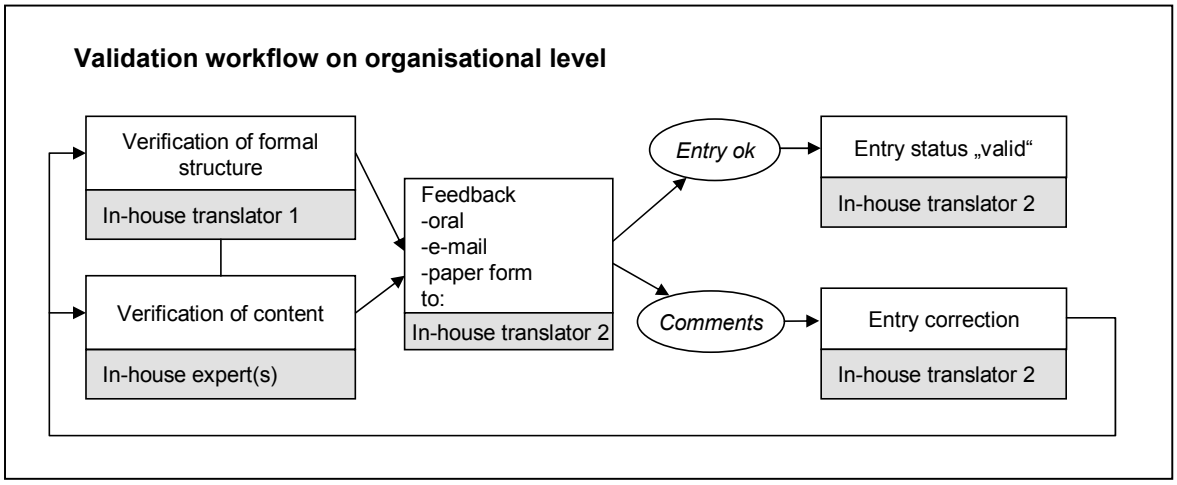

Figure 12.1. Validation workflow on organisational level

12.2 National level

12.2.1 Process requirements

On national level mono- or bilingual terminology is created for at least national-wide dissemination. Many experts from different institutions may take part in the terminology creation and validation workflow, which requires a high organizational effort. On this level the quality of the terminology has priority, time and financial restrictions being certainly less severe as they might be on organizational level.

Terminology developers are equipped with appropriate technical means — an electronic terminology management system, internet/intranet access, and possibly an integrated or external project management tool. The work shall be based on sound research principles, taking account the relevant international standards. So a well-considered and sophisticated database design based on ISO 12620 shall be taken for granted.

The examination of formal requirements consists of:

- Check for duplicates
- Integrity check — completion of all mandatory fields, no double completion of fields to be completed only once

- Format control — e. g. date formats, references, etc.
- Consistency check — functionality of cross references, correct form of the terms, e. g. singular for substantives, infinitive for verbs, case sensitivity, ISO language codes
- Spell check
- Grammatical check — correctness of the grammatical information to every term
- Classification control — correct assignment of the entry to a subject field of the chosen classification system.

The content check of every entry shall comprise the verification of

- Consistency of concept system
- Exhaustiveness of the terminology covering one subject field
- Choice of terms in every language according to the defined criteria, like linguistic correctness (i. e. ISO 704)
- Correctness of synonyms
- Exactness of definitions as regards content, comprehensibility of definitions (for wide dissemination), and formal requirements (writing rules defined for text fields, i. e. formulation of definitions according to ISO 704)
- Correctness of graphical representations
- Correctness of usage notes, temporal qualifiers, register, subject field, etc.
- Correctness of reliability codes assigned to terms.

12.2.2 Actors

For coordination purposes the competencies of all persons involved in the validation workflow have to be well defined. A terminologist shall be appointed project leader responsible for all validation processes. The creator of an entry must not check it by himself, but corrects the data according to the feedback of the reviewing persons; he might act as an updater as well.

The control of the formal requirements shall be performed by a translator or terminologist without subject field-specific knowledge, as far as it cannot be run automatically (cf. chapter *12.2.3 Feedback mechanisms*).

The content validation shall be carried out by experts of the respective subject field.

12.2.3 Feedback mechanisms

As far as possible, any information in the validation process shall be processed electronically. The project management tool might provide for a feedback routine, facilitating the coordination of the validation procedure. The entry mask for the terminological data shall contain validation fields and fields for additional comments. In order to ensure a transparent validation workflow which is crucial particularly for terminology databases updated and expanded on a regular basis, the project team shall work out a range of status codes at term level, according to ISO 12620.

The most effective way to validate the formal structure would be an automatic check run by a recording system allowing the final registration of a record only on condition that all formal requirements are met. In this case the creator of the entry directly clears the faults.

The subject field experts — either internal or external — are granted access to the databases via intranet or internet for the content-related check. If this is not possible, a feedback form shall be drafted and distributed to the experts either per e-mail or, should any electronic communication be impossible, in paper form. The experts either enter their comments directly into the system or return the feedback forms to the appointed responsible person, observing regular deadlines.

An electronic internet forum can be established, providing a platform for the discussion of controversial matters. If necessary, a personal meeting might be convoked by the project manager.

In case of a confirming answer, the terminological entry is validated by the creator or updater by changing its status code. Or else, the entries are modified according to the terminologists' and subject field experts' comments, and a new validation cycle starts.

12.3 International level

On international level, multilingual terminology for global dissemination is created by many experts from different countries. The coordination of the work and harmonization of the terminology is more labour-intensive and time-consuming as on national level, and high quality of the terminology has absolute priority. Nevertheless, the validation workflow is similar, so this chapter does not contain a detailed description. Please refer to the respective chapters under 7.10.2.

12.3.1 Process requirements

Cf. 12.2.1

12.3.2 Actors

The content validation shall be carried out by native speaking experts of the respective subject field. For the rest, please refer to chapter 7.10.2.2.

12.3.3 Feedback mechanisms

Cf. 12.2.3

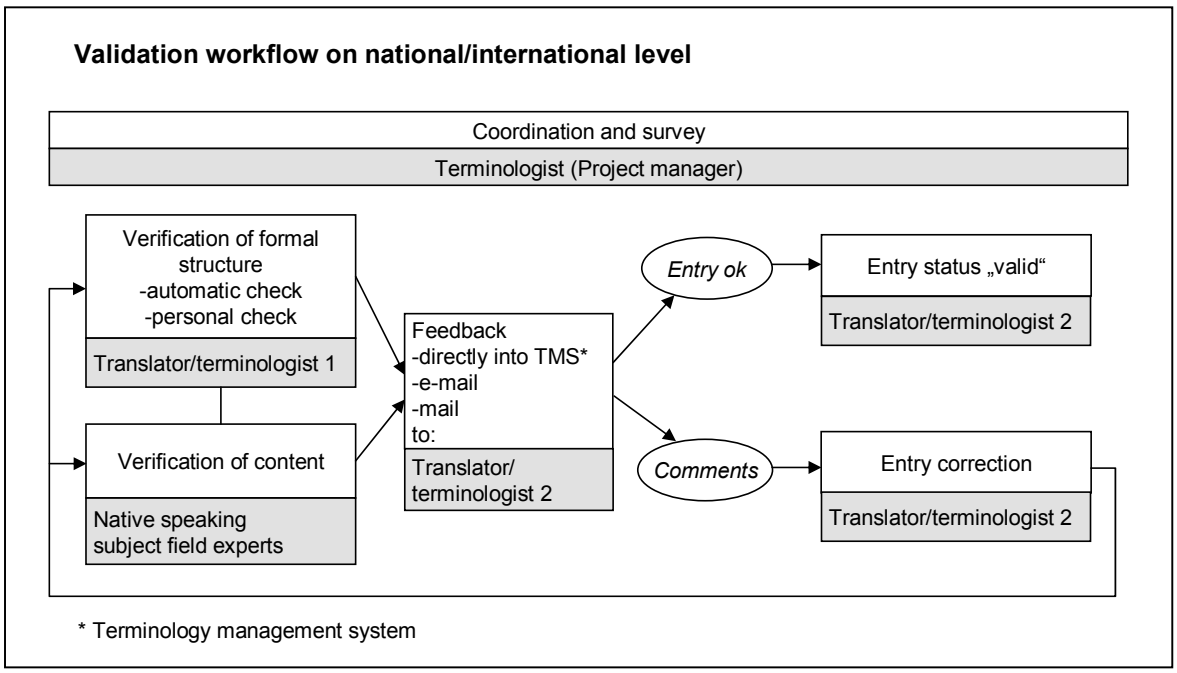

Figure 12.2. Validation workflow on national/international level

13 References

Deliverable 1.1, *Current Standards and Best Practices Assessment Report*, EuroTermBank project.

Appendix A

The below tables reflect assessments of best practice in terminology work carried out in the new EU member states.

The motivation for and the rationale behind this questionnaire was, in a general way, to collect information about assessments of best practice in terminology work in the new EU member states. The first task of workpackage 1 was to map and describe terminology work carried out in Estonia, Hungary, Latvia, Lithuania and Poland resulting in Deliverable 1.1. In order to be able to make general recommendations of best practice it was then considered necessary, in a controlled way, to assess the terminology work reported. These assessments were done by filling in evaluative scores reflecting which criteria (in a given terminology environment) that were regarded as being most important. The weight ranking system spanned from 1 to 3 in that 1 stood for lowest and 3 stood for the highest priority for a given criterion. For instance if *Possession of auxiliary tools in the terminological work* was considered as being very important in connection with 7.4 (Use of internet as a resource) then 3 should be plotted in the corresponding cell of the matrix.

The reference numbers on the horizontal axis correspond with the following topics within terminology work:
7.1 Overall workflow of terminology tasks
7.2 Classification Systems
7.3 Source identification
7.4 Use of internet as a resource
7.5 Compilation of terms and conceptual analysis
7.6 Data categories/7.7 Data Structure
7.8 Exchange format
7.9 Management of terminological entries
7.10 Validation workflow

Several observations can be made analyzing the filled in tables below. There is, however, one feature that is profoundly more significant compared to the other features expressed in the table. The correspondence between type of organization on the one hand and the criteria *quality in general terms* and *quality in terms of speed and efficiency* on the other hand forms a clearly distinctive feature. The table below illustrates this general observation.

No.	Type of organization	Quality of terms	Speed
1.	Organization	low to medium	high to medium
2.	National coordination	high to medium	low to medium
3.	International coordination	high	medium to high

Esterm by The Legal Language Centre

The descriptions of Esterm below are based on opinions of current and previous employees of The Legal Language Centre, the institution responsible for creating Esterm. The opinions were gathered over a time span of several years. The descriptions may therefore contain outdated opinions, inaccuracies or misinterpretations.

Framework: organization

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	3	2	2	3	1	2	2	2	2	2
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market	3				3	3	3	3	3	3
(III) quality in terms of broad coverage such as domain classification or information types		2				1	1			
(IV) Possession of auxiliary tools in the terminological work					1					
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists			3							
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3					1			3	3
(VII) Data category definition based on user needs						2	2	3		
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality									2	
(IX) Availability: How important is it that terminology is available to users outside the organisation?	3					3	3	3	3	
(X) Thorough validation routine → high reliability										1
(XI) Exchangeability with other term resources		3								

Creating terminological dictionaries for schools at the University of Tartu

The descriptions below are based on inside info from participants in a PHARE project, the aim of which is to create 12 terminology dictionaries for Russian-speaking pupils of age 14-16, with terms in Estonian and Russian, and definitions in Estonian.

Framework: organization

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	3	1	1	1	1	2	2		2	1
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market	2				3	2	2	3	2	2
(III) quality in terms of broad coverage such as domain classification or information types		1				1	1			
(IV) Possession of auxiliary tools in the terminological work							2	3		
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3		1		1	1				
(VI) Establishment of (well-prepared) procedures in the terminological workflow	2		1		1	1		3	2	1
(VII) Data category definition based on user needs	3					3	1	1	1	1
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality	3							3	2	3
(IX) Availability: How important is it that terminology is available to users outside the organisation?	3		3		3	3	3		3	
(X) Thorough validation routine → high reliability	2		1		1	1		3		1
(XI) Exchangeability with other term resources										

Framework: National coordinated level of Latvia

The main framework of term creation for TC of LAS is the national level. TC of LAS hosts 27 branch terminology commissions. The proposals of term creating – the request for creating a new term is addressed to one of these 27 branch commissions, which coin the very term. The process of terminology development of each branch is based on experience, using the system of existing terms, which were created by a set of definite principles. The TC of LAS is the official arbiter of the terms and the disseminator of them. Such a way of term coining envisages the necessity of a unified term system and definite criteria of terminology workflow.

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in terms of functionality e.g. complete filled in entries	3	3	3	3	3	2	2	2	2	3
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market	1				1					
(III) quality in terms of broad coverage such as domain classification or information types		3			3	2				
(IV) Possession of auxiliary tools in the terminological work		1			2					
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3	3								3
(VI) Establishment of (well-prepared) procedures in the terminological workflow	2									3
(VII) Data category definition based on user needs						1	1			
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality									1	
(IX) Availability: How important is it that terminology is available to users outside the organisation?	1							2		
(X) Thorough validation routine → high reliability	3									3
(XII) Exchangeability with other term resources								2		

Framework: National coordinated level of Lithuania

Terminology work at the Institute of Lithuanian Language was started in 1941. A separate Department of Terminology was founded in 1991 and it was reorganized into the Centre of Terminology in 2003. Dr Albina Auksoriūtė is Head of the Centre and the staff of twelve includes four people with doctorates

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in terms of functionality e.g. complete filled in entries	2		2	1	2	3				
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market	2				2	2				1
(III) quality in terms of broad coverage such as domain classification or information types		3				3	3			
(IV) Possession of auxiliary tools in the terminological work	1			1	1	1		2		
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3				2	2				3
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3				2	3			3	3
(VII) Data category definition based on user needs		3				3				
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality									2	
(IX) Availability: How important is it that terminology is available to users outside the organisation?								1		
(X) Thorough validation routine → high reliability	3									3
(XII) Exchangeability with other term resources		1						2		

Framework: organizations in Poland

Contribution by OPI based on PKN, UKIE and PolTerm experience

Topics according to Del 1.1 chapter 7 / Parameters/criteria	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	3	2	3	2	2	2	2	2	3	3
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market		2								
(III) quality in terms of broad coverage such as domain classification or information types		1				2	2			
(IV) Possession of auxiliary tools in the terminological work				1	1					
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3									3
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3					2			3	3
(VII) Data category definition based on user needs						1	1			
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality									2	
(IX) Availability: How important is it that terminology is available to users outside the organisation?						1	1	3		
(X) Thorough validation routine → high reliability	3									3
(XI) Exchangeability with other term resources		1								

EUJog by the Hungarian Ministry of Justice, MorphoLogic, and SZAK Publishers

In Hungary, there are no permanent national organizations with the purpose of creating terminology resources.

Details below are listed according to MorphoLogic's experience with the practices of the Hungarian Ministry of Justice, also taking into account opinions from SZAK Publishers, the company that rendered the EUJog termbank into a printed dictionary.

Framework: national organisation in connection with local organisations

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	3	2	2		2	3	1		2	2
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market					1	1				
(III) quality in terms of broad coverage such as domain classification or information types		2				2				
(IV) Possession of auxiliary tools in the terminological work				2	2	2				
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3	2	3	2		2				
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3			3	2	2		2	3	3
(VII) Data category definition based on user needs		2				2				
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality				2	3	3	3	3	3	
(IX) Availability: How important is it that terminology is available to users outside the organisation?	3	2	3				1	3		
(X) Thorough validation routine → high reliability	3		2		3	2	2	2		3
(XI) Exchangeability with other term resources	3	2	3			2	1	3		

Creating terminological glossaries for book translation purposes

The descriptions below reflect the practices applied by SZAK Publishers, a private organization using a systematic terminology workflow to prepare pre-compiled glossaries for book translation processes, where the translators receive a complete term base before starting their work.

Note: SZAK Publishers is one of the founding organizations of the Hungarian Terminology Council, and it is the only private organization to delegate one of the Council's vice presidents.

Framework: local organisation

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	1	1	2	2	1	1	2	1	3	2
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market	3				3	3		3	3	3
(III) quality in terms of broad coverage such as domain classification or information types										
(IV) Possession of auxiliary tools in the terminological work	2			2				3	2	
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3		3	3	2	3				3
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3		3		2	2	3	3	3	3
(VII) Data category definition based on user needs						1				
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality	3		3		3	3	3	3	2	
(IX) Availability: How important is it that terminology is available to users outside the organisation?	1						3	3	2	
(X) Thorough validation routine → high reliability	3		1	2	1	1		3		3
(XI) Exchangeability with other term resources			2		2	1	1	2	2	2

Contribution from CST

The below descriptions of IATE are based on CST's experiences in the IATE project. The descriptions may therefore contain inaccuracies or misinterpretations.

Framework: International coordination

Topics according to Del 1.1 chapter 7 / **Parameters/criteria**	7.1	7.2	7.3	7.4	7.5	7.6	7.7	7.8	7.9	7.10
(I) quality in general terms	3	2	3	1-3♦	1-3♦	3	3	3	3	3
(II) quality in terms of speed and efficiency e.g. having the strategy of getting fast-to-market										
(III) quality in terms of broad coverage such as domain classification or information types		2				3	3			
(IV) Possession of auxiliary tools in the terminological work		3								
(V) Accessibility to expert knowledge, e.g. domain experts/terminologists	3	3								3
(VI) Establishment of (well-prepared) procedures in the terminological workflow	3					2(WritingRules)			3	3
(VII) Data category definition based on user needs						3	3			
(VIII) Technical complexity/design preferences: e.g.. making the system easy to maintain and fast to update with the loss of advanced functionality									2	
(IX) Availability: How important is it that terminology is available to users outside the organisation?						1	1	3		
(X) Thorough validation routine → high reliability	3									3
(XI) Exchangeability with other term resources		3				3	3	3		

♦ The IATE termbase entries is a union set, which is the result of all involved partners' ongoing contributions. This IATE group is very heterogeneous in terms of size and available resources. This fact is reflected in their different approaches to the topics 'Use of internet as a resource' and 'Compilation of terms and conceptual analysis' explaining why the score is a scale rather than fixed variable.