TOWARDS CONSOLIDATION OF EUROPEAN TERMINOLOGY RESOURCES

Experience and Recommendations from EuroTermBank Project

TOWARDS CONSOLIDATION OF EUROPEAN TERMINOLOGY RESOURCES

Experience and Recommendations from EuroTermBank Project

Tilde, Riga

Towards Consolidation of European Terminology Resources
Experience and Recommendations from EuroTermBank Project

Edited by: Signe Rirdance, Andrejs Vasiļjevs

Authors and contributors: Albina Auksoriūtė, Imants Belogrīvs, Audronė Bielevičienė, Ainars Blaudums, Jānis Bordāns, Tomasz Borkowski, Juris Borzovs, Anna Braasch, Eduards Cauna, Anita Dravniece, Nijolė Dudlauskienė, Christian Galinski, Lina Henriksen, Heiki-Jaan Kaalep, Andris Kalniņš, Danuta Kierzkowska, Helle Kilgi, Balázs Kis, Andris Liedskalniņš, Bente Maegaard, Asta Mitkevičienė, Sussi Olsen, Claus Povlsen, Uldis Priede, Gábor Prószéky, Ivars Puksts, Marin Raguz, Inke Raupach, Rauno Salupere, Klaus-Dirk Schmitz, Krystyna Siwek, Raivis Skadiņš, Valentīna Skujiņa, Robertas Stunžinas, György Tardy, Anu Uritam, Andrejs Vasiļjevs, Jūratė Zabielaitė

Acknowledgements: The EuroTermBank Consortium would like to extend their gratitude to everybody who was involved in reaching the ambitious goals of this project, including the EuroTermBank content providers and authors, survey participants, European Commission officers, colleagues, friends, and family.

The EuroTermBank project partners:

- Centre for Language Technology at University of Copenhagen, Denmark
- Information Processing Centre, Poland
- Institute for Information Management at Cologne University of Applied Science, Germany
- Institute of Lithuanian Language, Lithuania
- MorphoLogic, Hungary
- Terminology Commission of the Latvian Academy of Sciences, Latvia
- Tilde, Latvia (project coordinator)
- University of Tartu, Estonia

This publication is co-financed by the European Commission eContent programme.

Copyright © EuroTermBank Consortium, 2006
All rights reserved. No part of the contents of this book may be reproduced or transmitted in any form or by any means without the written permission of the publisher.
Product and company names mentioned herein may be the trademarks of their respective owners.

Published by: Tilde, Vienibas gatve 75A, Riga, LV-1004, Latvia
Printed in Riga, Latvia

ISBN 9984-9133-4-1

CONTENTS

Preface 7

Introduction 9

EuroTermBank project overview 11

Chapter 1: Methodology recommendations in terminology management 13

1.1 Approaching terminology best practice 13
1.1.1 Goals and conditions of terminology management 14
1.1.2 International, national and local scenarios 15

1.2 Workflow of terminology tasks 16
1.2.1 International scenario 16
1.2.2 National scenario 17
1.2.3 Local scenario 18
1.2.4 Translation workflow in Estonian Legal Language Centre: a case study 20
1.2.5 Terminology creation workflow in Terminology Commission of Latvian Academy of Sciences: a case study 20

1.3 Classification systems 21
1.3.1 Lenoch universal classification system 21
1.3.2 Eurovoc 23
1.3.3 Comparison and use of terminology classification systems 24
1.3.4 Local scenario 24
1.3.5 National scenario 24
1.3.6 International scenario 24

1.4 Concept analysis 24
1.4.1 Local scenario 26
1.4.2 National scenario 27
1.4.3 International scenario 28

1.5 Data structure and data categories 29
1.5.1 Local scenario 29
1.5.2 National scenario 30
1.5.3 International scenario 31
1.5.4 IATE data categories: a case study 32
1.5.5 PolTerm data categories: a case study 33

1.6 Exchange format 33
1.6.1 Local scenario 34
1.6.2 National scenario 34
1.6.3 International scenario 34

1.7 Validation workflow 35
1.7.1 Local scenario 35
1.7.2 National scenario 36
1.7.3 International scenario 38

1.7.4 Validity of term entries in IATE: a case study 39
1.7.5 Validity of term entries in PolTerm: a case study 39
1.7.6 Validity of term entries in Estonian Legal Language Centre: a case study 39
1.7.7 IATE writing rules: a case study 40

Chapter 2: Terminology infrastructure and standardization 43

2.1 State of terminology management in selected new EU member countries 43
2.1.1 Estonia 43
2.1.2 Hungary 45
2.1.3 Latvia 45
2.1.4 Lithuania 47
2.1.5 Poland 48

2.2 Termbanks, bodies and networks 49
2.2.1 IATE: Inter-Active Terminology for Europe 49
2.2.2 Infoterm: a network of terminology centres 50
2.2.3 TERMIUM: the Canadian termbank 51
2.2.4 TSK: the Finnish Terminology Centre 51
2.2.5 Glimpses of terminology work in global companies 52
2.2.6 Selected projects related to terminology field 53

2.3 Terminology tools 56
2.3.1 Terminology management tools 56
2.3.2 Term extraction tools 57

2.4 Terminology standardization 58
2.4.1 ISO standards in the terminology field 59
2.4.2 ISO 704: Terminology work – Principles and methods 60
2.4.3 ISO 860:1996 Terminology work – Harmonization of concepts and terms 60
2.4.4 ISO 10241: International terminology standards –Preparation and layout 61
2.4.5 ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange 62
2.4.6 ISO 12620:1999 – Computer applications in terminology – Data categories 64
2.4.7 ISO 16642:2003 – Computer applications in terminology – Terminological markup framework (TMF) 66
2.4.8 Terminology standardization in Germany: a case study 69

Chapter 3: Legal framework for terminology databases 71

3.1 Overview on copyright questions 71
3.1.1 Copyright in terminology 72
3.1.2 Groups of copyright holders 74

3.2 EU legislation on copyright and related rights 75
3.2.1 Directive 96/9/EC on the legal protection of databases 76
3.2.2 Directive 2001/29/EC on the harmonization of certain aspects of copyright and related rights in the information society 77
3.2.3 Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market 78

3.2.4 Directive 2004/48 EC on the enforcement of intellectual property rights 79
3.3 Code of Good Practice for Copyright in Terminology by Infoterm 79
Chapter 4: Evaluation and description of terminology resources in EuroTermBank 83
4.1 Evaluation and selection of resources 83
4.1.1 Considerations for evaluation of terminology resources 83
4.1.2 Guidelines and criteria for the evaluation of terminology resources 84
4.2 Terminology resource description (TeDIF) 86
Chapter 5: The EuroTermBank portal 89
5.1 System overview 89
5.2 Usage scenarios 90
5.3 EuroTermBank services 91
5.3.1 Search 91
5.3.2 Editing 92
5.3.3 Import 92
5.3.4 Export 92
5.3.5 Access for 3rd party software 92
5.3.6 EuroTermBank discussions services 92
5.4 System architecture 93
5.4.1 Overview 93
5.4.2 Types of users 94
5.4.3 User interface layer 94
5.4.4 Business logic layer 95
5.4.5 Database 95
5.4.6 External terminology databases 95
5.5 Data structure and exchange format 96
5.5.1 Data structure 96
5.5.2 Data categories 97
5.5.3 Data exchange format 99
Appendices 101
Appendix A. Data structure 101
Appendix B. The ETB data exchange format 107
Appendix C. References from EuroTermBank project documentation 117
Literature 117
Standards and EU directives 121
Internet resources 122

PREFACE

This publication summarizes the experiences and findings of the EuroTermBank project, part of the European Union eContent program. It is aimed at individuals and organizations interested and involved in all aspects of terminology management. The following major areas are covered in this book:

- Methodology recommendations in terminology management, based on best practices
- Insights into the national and international terminology infrastructure, with a focus on selected "new" countries of the EU
- A summary on ISO terminology standards
- Insights into the legal framework of terminology work
- Description of the EuroTermBank portal (http://www.eurotermbank.com)

This publication is not an attempt to comprehensively cover the field of terminology; rather, it provides glimpses and insights into the most valuable findings and results of this project. The hope is, however, that these can be of value as a reference point for researchers and students, translators and technical writers, software developers and many other workers and parties in the truly exciting area of terminology.

We thank the project partners – Tilde, the project coordinator (Latvia), Institute for Information Management at Cologne University of Applied Science (Germany), Centre for Language Technology at University of Copenhagen (Denmark), Institute of Lithuanian Language (Lithuania), Terminology Commission of Latvian Academy of Sciences (Latvia), MorphoLogic (Hungary), University of Tartu (Estonia), and Information Processing Centre (Poland), for their dedication and hard work.

INTRODUCTION

Consistent, harmonized and easily accessible terminology is an extremely important stronghold for ensuring true multilingualism in the European Union and throughout the world. From EU legislation and trade to the needs and mobility of every EU citizen, terminology is the key for easy, fast and reliable communications. The rapid path of changes in many technological and economical areas leads to ever growing introduction of new concepts and terms to describe them. Globalization from the one side and growing language awareness from the other side dictates the need to consolidate different national terminology resources, harmonize international terminology, and provide online access to reliable multilingual terminology.

Availability of comprehensive and accessible terminology resources is a growing requirement for economical and social development. As we face continuous growth of Internet penetration, centralization of access to multilingual terminological resources becomes crucial.

The major terminology developers include public institutions, universities, technical societies as well as representatives of the private sector. Although a significant number of such institutions do exist, only a few of them produce resources that are exchangeable and/or marketable.

In most of the countries there is a lack of coordination between institutions dealing with terminological activities. This often results in useless efforts or duplicate results. Terminologists and subject specialists have little contact with their colleagues working on similar subject areas. Across subject fields and in different sectors potential users of terminology are often not even aware what resources are available.

The reason for this lack of communication is the general fragmentation of creation and distribution mechanisms on the institutional, sector/industry and national levels. Term banks tend to be small in size, mostly highly specialized, difficult to access. These difficulties are amplified by considerations of confidentiality, institutional restrictions and legal uncertainty about copyright status of certain terminology resources.

As a result, there is a lack of terminology resources and the existing resources are not adequately reutilized. Quality of available terminological collections varies widely and is inadequate in many cases. International standards are not always used in terminology development and may sometimes be even not known to the people directly involved.

Differences in methodological approach, structuring and formatting in creation of terminology data on institutional and national level are an important terminology quality issue. Different terminology development procedures exist in different institutions and countries and the variety of poorly coordinated public and private terminological organizations, sometimes creating parallel inconsistent versions of the same terms.

These problems are particularly acute in new European Union member countries that have undergone rapid social and economic transformations and are in urgent need to integrate their terminology development with the rest of the EU and global economy. The overall situation in terminology area is characterized by many gaps and problems. New EU member countries face the issue of terminology resource fragmentation across different institutions, inconsistency and lack of coordination in terminology development, as well as structural and technical incompatibility.

A lot of terminology data are available only in the form of printed dictionaries and bulletins or stored in card files. The transformation from a centralized terminology development during Soviet time with the focus on Russian language to the requirements of market economies is still not fully completed. It has lead to lack of coordination between institutions involved in terminology development, inconsistency and poor quality of terminology data, insufficient mechanisms for dissemination of new terminology.

Currently there are several important multilingual terminology resources for the languages of the "old" EU countries – Eurodicautom, which holds terminology of European Commission in eleven languages, TIS (Council of Ministers), Euterpe (European Parliament) and IATE project for single database for EU institutions. Only a small part of terminology entries in these databases cover languages of new EU member countries. These resources include mostly official EU terminology, leaving aside many other areas, and a lot of public/private terminology resources are not networked and are difficult to access and use for wider public.

Rapid development and dissemination of new terms are especially important for smaller languages. There are several initiatives underway to create national terminological databases. For example, the Latvian project partners, the Terminology Commission of the Academy of Sciences and Tilde have participated in an initiative creating an online database. However, these initiatives are limited due to lack of resources and they focus only on national terminology.

An important initiative to address these weaknesses and problems is the EuroTermBank project with the goal to collect, harmonize and disseminate dispersed terminology resources through an online terminology data bank.

EUROTERMBANK PROJECT OVERVIEW

The EuroTermBank project focuses on harmonization and consolidation of terminology work in new EU member states, transferring experience from existing European Union terminology networks and accumulating competencies and efforts of the accessed countries.

The project has resulted in a centralized online terminology bank for languages of new EU member countries interlinked to other terminology banks and resources. Although EuroTermBank is addressed directly towards Estonia, Hungary, Latvia, Lithuania, and Poland, the project is open to other EU member states and interested countries and organizations outside EU.

It enables the exchange of terminology data with existing national and EU terminology databases by establishing cooperative relationships, aligning methodologies and standards, designing and implementing data exchange mechanisms and procedures. Through harmonization, collection and dissemination of public terminology resources, EuroTermBank strongly facilitates enhancement of public sector information and strengthens the linguistic infrastructure in the new EU member countries.

The main goal of the EuroTermBank project is to contribute to improvement of the terminology infrastructure in the new EU member countries. This aim is accomplished by establishing terminology networks and by collection and harmonization of existing terminology resources resulting in an implementation of a centralized online term base.

EuroTermBank project is launched by 8 partners from 7 European Union countries – Germany, Denmark, Latvia, Lithuania, Estonia, Poland and Hungary. The project partners are Tilde (Latvia), Institute for Information Management at Cologne University of Applied Science (Germany), Centre for Language Technology at University of Copenhagen (Denmark), Institute of Lithuanian Language (Lithuania), Terminology Commission of Latvian Academy of Sciences (Latvia), MorphoLogic (Hungary), University of Tartu (Estonia), Information Processing Centre (Poland).

The project is part of the European Union eContent program which aims to facilitate the production, use and distribution of European digital content and promote linguistic and cultural diversity on the global networks.

The project focuses on the following major objectives:

- Development of methodology for harmonization of terminology processes in new EU member countries and for ensuring compatibility of terminological resources for data interchange and resource sharing;
- Creation of a network of terminology-related institutions and organizations (creators and holders of terminology resources) on both national and multinational levels to facilitate institutional cooperation and harmonization, consolidation and dissemination of terminological resources;
- Design, development and implementation of a web-based terminology data bank to provide easy access to centralized terminology resources;
- Consolidation of terminology content from different sources and owners for creation of national terminological databases and further integration into the EuroTermBank database or their interlinking;
- Achieving sustainability of the project results.

Development, population and maintenance of a web-based terminology data bank constitute the major tangible outcome of the project. The data bank works on a two-tier principle – as a central database and as an interlink node or a gateway to other national and international terminology banks.

Data exchange mechanisms are developed to establish term import, export and exchange with other terminology databases. The results of project activities in harmonization and standardization form a unified basis for term exchange technologies.

Terminology content is of highest importance in the project. In brief, the ultimate objective is to integrate all available terminology resources (not only from project partner countries) into the central EuroTermBank database or interlink them via EuroTermBank as a central gateway and a single point of service.

In general, terminology content is available in two forms – electronic and hardcopy (i.e. card files). Regardless of the type, the content requires additional processing to be ready for integration into EuroTermBank.

The methodology developed in the EuroTermBank project serves as the basis for content processing. The content passes several stages before integration into the database, including selection, prioritization, modification, and digitalization (for non-digital format).

The outcome of this process is a reliable multilingual terminology resource, networked with other existing national and international resources available for users over the global network. An example of a national terminology database interconnected with EuroTermBank is the online databank of Latvian official terminology. An example of an international databank that could greatly benefit from interlinking with EuroTermBank is IATE, the termbase of European Union official institutions.

Selection principles defined within the project context ensure that the pool of existing terminology resources collected in EuroTermBank meets the quality criteria reflecting the needs and demands of the users. Specification of the term base is prepared with a view to international data exchange standards facilitating implementation of exchange mechanisms for term data from other EU terminology resources.

The overall project plan contains a number of tasks. First, an inventory of international standards and best practices in terminology work and term management in involved new EU member countries was established and recommendations for best methodology were prepared. With a view to these recommendations and conducted surveys of user needs and requirements, the specification of the system and the database platform was created. This specification contains a description of the overall architecture and design, data categories and structure, system functional specification and interface description.

After the implementation phase including a pilot trial and the standard software evaluation methodology, the final step in the project plan was the validation phase, where the implemented functionality was validated against the system specification.

The project has resulted in a centralized web-based terminology bank for languages of the new EU member countries interlinked to other terminology banks and resources.

CHAPTER 1

METHODOLOGY RECOMMENDATIONS IN TERMINOLOGY MANAGEMENT

This chapter provides recommendations for best practice within the field of terminology management. It is based on the assessment of 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. This assessment was carried out in the framework of EuroTermBank project in order to extract best practice within all the aspects of terminology methodologies.

1.1 Approaching terminology best practice

An important aim of the EuroTermBank project was to specify best practice within most areas of terminology work, ranging from the use of terminology tools and classification systems to concept analysis and design of termbases. Best practice of terminology management is however dependent on circumstances and on the particular context of the terminology work. Therefore several terminology scenarios were established in order to represent schematic frameworks of terminology work.

In the EuroTermBank project context, best practice in terminology work is based on existing international standards and on a survey of 'real-life' terminology work as it is conducted in the new as well as the old EU member countries. Among the terminology resources that were investigated are for example the state regulated or coordinated terminology collections of the new EU member countries and the IATE terminology cooperation of the old EU countries.

International standards have been used as a starting point for development of best practice. However, standards are very general and describe recommendations in a vacuum disconnected from specific goals and preferences and also disconnected from the set of conditions that apply in a given context. By conditions we refer to the premises or state of things that cannot be changed easily or at all. For example, a condition might be that all language professionals of a particular organization do not have access to the internet or to terminology tools. Therefore it was necessary not only to investigate how terminology work is actually carried out in different settings, but also to investigate the conditions and goals of the particular terminology settings.

In this chapter we will describe the conditions and goals that have been identified during the project, describe the different terminology scenarios that we have based our work on and give examples of best practice within several different aspects of terminology work. Some of the described aspects of terminology work are supported and complemented by case studies extracted from the survey of real-life terminology methodologies.

1.1.1 Goals and conditions of terminology management

A survey of terminology settings in the new as well as the old EU states was conducted, to identify terminology management goals and conditions. In cooperation with terminology resource owners, project partners prepared an assessment of the influence of each condition and the importance of each goal. The aim was, as a first step towards establishment of a number of fixed scenarios with best practice descriptions for each terminology task, to identify sets of goals and conditions that typically co-exist.

The following table shows the goals identified by the resource owners and considered as having a profound impact on terminology methodologies.

Goal	Explanation
High quality in general terms	High quality in general terms means that terminology work is based on sound research principles; consistent, exhaustive, non-ambiguous, broadly accepted etc.
Harmonization	Harmonization concerns concepts as well as terms and a harmonization process involves comparison of concept systems: relations between concepts, number of concepts, depth of structure, deletion of duplicate concepts etc. leading to construction of a new harmonized concept system. In many contexts an inherent part of 'high quality', but harmonization is only relevant in some scenarios.
Exchangeability	Exchange of data between term resources using standard approved exchange methodologies
Availability	Terminology must be available to external users, i.e. to users outside the particular organization.
Speed and up-to-dateness	Speed of terminology work and data that are always up-to-date are considered particularly important

The following table shows the conditions identified by the resource owners and considered as having a profound impact on terminology methodologies.

Goals/Scenarios	International	National	Local
Access to tools	Access to terminology tools	Access/no access to terminology tools	Access/no access to terminology tools
Professional representation	All types of language professionals represented	All types of language professionals represented	Terminologists and subject field experts often not part of terminology developer team
Financing	Adequate financial support	Adequate financial support	Often a tight budget
Number of languages	Multilingual	Mono- or bilingual	Bi- or multilingual
Domain coverage	Broad domain coverage	Broad domain coverage	Focused domain coverage
Area of activity	Coordination (translation)	Coordination (regulation, translation)	Translation

1.1.2 International, national and local scenarios

The scenarios that were identified within the ETB project research are based on the distinction between international, national and local terminology settings. In this context, a terminology scenario means a schematic framework of terminology work that is based on a certain set of conditions and goals.

The international scenario (or level) is concerned with coordination and management of multilingual terminology work in a well-organized infrastructure and primarily concerns approval or dismissal and harmonization of terms that are coined at the national and local levels.

The main distinctive element is that terminology work at the national level is usually mono- or bilingual. The main activities in the national scenario (or level) are similar to those of an international scenario. Another difference is that organizations belonging to the national framework may have national regulatory obligations.

The local scenario (or level) covers organizations that do not belong in an international or national framework and concerns terminology work that originates from translation and creation of documents. This type of terminology work involves identification of national language terms in relevant documents, terminology glossaries, available terminology bases and possibly in relevant literature of the domain. When a national language term cannot be identified, new terms are coined. Characteristic features of a local framework are that terminology work is usually limited to one or a few closely related domains, harmonization often does not play a significant role, and restricted budgets and tight time frames are more likely than in national or international frameworks.

These three scenarios (levels) represent only schematic frameworks of terminology work. Requirements, aims and circumstances can differ somewhat even within one framework. Therefore best practice described for one scenario may also in some cases be applicable for an organization that would in this context belong to another scenario. Besides, some additional factors may play a role, for example, the core business and the size of the particular organization. The impact of these factors however is difficult to measure.

The following table shows typical goals in the international, national and local scenarios.

Goals/Scenarios	International	National	Local
Quality	High quality in general terms	High quality in general terms	Tight time frames coexist with (and put limitations on requirements for) high quality
Harmonization	Harmonization is high priority	Harmonization is high priority	Harmonization is not a priority
Exchangeability	Exchangeability is high priority	Exchangeability is high priority/is sometimes not a priority (recommended as high priority)	Exchangeability is often not a priority (recommended as high priority)
Availability	Availability is high priority	Availability is high priority	Availability is not a priority

The following table shows typical conditions in the international, national and local scenarios.

Goals/Scenarios	International	National	Local
Access to tools	Access to terminology tools	Access/no access to terminology tools	Access/no access to terminology tools
Professional representation	All types of language professionals represented	All types of language professionals represented	Terminologists and subject field experts often not part of terminology developer team
Financing	Adequate financial support	Adequate financial support	Often a tight budget
Number of languages	Multilingual	Mono- or bilingual	Bi- or multilingual
Domain coverage	Broad domain coverage	Broad domain coverage	Focused domain coverage
Area of activity	Coordination (translation)	Coordination (regulation, translation)	Translation

Terminology work involves many types of tasks, and most of these tasks have been dealt with in the EuroTermBank project with a view to extracting best practice. For this publication, some of the most significant terminology tasks have been selected and best practice is described for each scenario: the local, the national and the international scenarios.

1.2 Workflow of terminology tasks

Terminology work is performed differently in different countries and settings. How the work is carried out depends on the organizations, institutions or bodies engaged in the process as each of these defines individual goals and objectives. Terminology work is usually arranged in accordance with specific goals, and the activities performed may depend on the composition of the staff managing and developing the terms. Consequently, the number and types of stages of terminology work differ as well.

Below, we summarize the tasks performed and the working process arrangements, also called the workflow, at international, national and local levels or scenarios.

1.2.1 International scenario

Many of the activities at this level are related to international co-operation and standardization aiming at a production of comprehensive, high-quality and reliable terminology stored in term bases. Terminological standardization concerns principles, methods and applications related to terminology and standardization of terms as well. An important task is dissemination of the developed standards and terminology.

The workflow is usually based on a well-defined set of tasks and procedures, which means that all terminological standardization is strictly regulated, including rules for the workflow. An illustrative example is the terminology standard development process at TC 37 committee of ISO, which comprises the following six steps:

- Proposal stage
- Preparatory stage
- Committee stage
- Enquire stage
- Approval stage
- Publication stage

Although the character of terminology work differs from one organization to another, some common, typical features are identified; these provide a basis for our general recommendations.

Terminology collections at the international level are multilingual; a very important feature which does not apply to other levels. The overall objective is to perform high quality work. In general terms it is achieved by means of developing appropriate procedures, involving highly qualified staff.

On the basis of the terminology work reported for the international level, the following general recommendations for best practice can be summarized:

- The staff should comprise domain experts, different types of language professionals (e.g. translators and interpreters), domain experts and terminologists (from various countries).
- Terminology tools should be integrated into translation and office automation environments, easy access to internet and on-line communication are preconditions as well.
- The work at international level should optimally include a coordination of terminology work between the different countries and institutions involved; it is also necessary to ensure data exchangeability. A further essential task is terminology harmonization between domains.
- Well-prepared procedures in the terminological workflow and strict validation procedures should be employed in order to ensure the achievement of the overall high-quality goal.
- The developed terminology should be available to users outside the international cooperation through online access, either free of charge or for a fee.

1.2.2 National scenario

In spite of the differences in activities, the basic tasks are common to most institutions and organizations (e.g. national councils and commissions), and they are the following:

- terminology and language planning
- development of integrated terminology systems based on international principles
- national standardization and approval of terms
- maintenance of national terminology
- coordination of terminological work in state institutions, standardization departments, translation centres and other organizations.

In some organizations, the activities carried out comprise not only terminology work but also some general language related tasks, such as language normalization and standardization, coordination of language policy and terminology work.

The most important task at the national level is harmonization between domains, because of the diversity of experts and institutions involved in the terminology creation workflow. Further, terminology work at this level may be interrelated with work carried out at the local level, as extraction and creation of terms and definitions are often carried out in the local scenario, while expertise and approbation of terms and definitions are carried out at the national level. This fact is of relevance for the organization of the workflow.

Terminological work at the national level is mainly monolingual or bilingual. The following can be concluded on best practice and general recommendations.

Access to and use of terminology tools and the internet should be easy in order to support an efficient cooperation between domain specialists. National term banks usually serve as the main terminology tool of actors in terminology development.

Appropriate time and financial conditions should be ensured. (Although the state support is sometimes inadequate, at the national level these conditions are usually not as limited as at the local level.)

The developed term collections are created for national dissemination and should be accessible to the national user community. National bodies can also ensure a consistent usage of normalized terms, create information systems, etc.

1.2.3 Local scenario

Local terminology work is performed by organizations such as translation agencies, documentation centres, research institutes, etc. They of course organize their work differently and apply different work practices. Their workflow is highly dependent on the type of the particular terminology setting.

The organizations differ not only in terms of the type of their work but also in terms of the staff employed. Translation bureaus usually employ translators, language specialists; sometimes experts of different domains are involved as well, while the involvement of terminologists is quite rare.

The activities of research institutions usually involve linguist terminologists only. Occasionally, in connection with general terminology projects are relevant specialists and terminologists invited in their capacities of consultants or experts.

The entities working at a local level have one central feature in common: they manage and develop terminology of one or of a few, usually related, domains. Therefore, harmonization of concepts and terms between domains is usually not relevant. This is one of the main differences between the local scenario on the one hand and the national and international scenarios on the other.

As an illustration, the following two examples are given:

- Organizations translating EU legislation. First, translators search for the corresponding terms needed in their national language in related documents, terminology glossaries and other available terminology sources of the domain. If the term cannot be found, 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.
- Research institutions deal mainly with theoretical research on terminology, but terminologists at these institutions also do practical work on normalization of terminology and work together with specialists of various domains. Terminologists give recommendations to specialists of a particular domain in regard to naming the concepts in the most appropriate way.

Terminological work at the local level is mainly defined by the user's needs e.g. translation or localization of documents, etc. and their working conditions, e.g. the framework of research projects. Speed is a basic requirement in terminology management and development tasks, e.g. the translation bureaus have to observe the deadlines.

On the basis of reports contributed by various local actors to the present study, the following can be concluded on present practice.

The types of terminology sources and tools available to the terminology developers are very important. The new European member states have quite a number of printed terminology dictionaries in different domains; electronic terminology dictionaries gradually emerge as well. On the other hand, terminology developers at the local level often do not have access to the international terminology banks and internet-based databases, whereas a similar problem is not faced in national and international scenarios.

As regards the staff employed, terminologists are only rarely involved at the local level, thus the translators' professional skills including knowledge of the native language and main principles of terminology are of crucial importance.

The terminology developed in the local scenario is, as a rule, not approved and endorsed by any competent national organization, while such practice is common to the national level. Some general recommendations:

- It is very important to perform terminology management and development tasks at the local level effectively, ensuring at the same time reliable and high-quality terminology.
- The involvement of high-profile translators and different specialists with thorough knowledge of both the language and the key principles of terminology work is highly recommended.
- The staff should be able to consult reliable terminology sources. In addition, the work quality significantly improves if the staff includes terminologists or terminologists are being constantly consulted. (Quite a few entities on the organizational level apply such practice.)
- A good practice comprises also the submission of newly coined terms to a relevant national body for approval.

A further relevant issue is related to the quality of terms developed and introduced by the organizations. At the national level, consistent, correct and adequate terms are introduced, which are harmonized with the terminology in other domains too. Although such a complex 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 local level as well.

1.2.4 Translation workflow in Estonian Legal Language Centre: a case study

The Estonian Legal Language Centre is a public organization under the Ministry of Justice and the purpose of this organization is to meet the legislative translation and terminology needs of the Estonian Government. In brief, the workflow in connection with translation of documents contains the following steps:

1. A new document to be translated is passed to a terminologist who
    - identifies terms
    - checks existence of term translations
    - if the term exists, adds the document ID to the term entry
    - if there is no equivalent term, creates the missing term
    - if the text is highly specialised, recommends translation by an expert contractor
    - stores the research materials in a paper folder and the result – in the term database
2. The document with marked-up terminology is passed to a translator.
3. The translator may contact the terminologist for clarifications or improvements of terminology.
4. The terminologist updates term entries in the database, if necessary.

The terminologists and translators use Trados Translator's Workbench and MultiTerm in their work.

1.2.5 Terminology creation workflow in Terminology Commission of Latvian Academy of Sciences: a case study

The Terminology Commission of the Latvian Academy of Sciences uses the following main steps in terminology creation:

- The initiators of a term may be different organizations and individuals dealing with all kinds of social, legal, technological, medical, economic activities; translators and education managers, banking officers, newspaper writers and publication editors, standard developers and even students and front-end users of different kind of devices.
- The request for a new term is addressed to the corresponding branch of a subcommission of the Terminology Commission of the Latvian Academy of Sciences, which at its meeting analyses the requested term and its explanations. Explanations of the term may be found in online dictionaries, and they may have different degree of detail. The task of subcommission members is to select the essential features and propose an adequate Latvian equivalent of the term. The TC of LAS hosts 26 subcommissions, each responsible for its branch of terminology.

- After the harmonization of terms within a particular branch, they have to be harmonized with terms of related branches and with the whole Latvian lexical system. The TC of LAS is responsible for this task. TC of LAS is also the main arbiter to decide about borrowing or not borrowing English terms from EU documentation and creating new Latvian terms if the corresponding term could not be found in the present vocabulary.
- The terms approved by TC of LAS are published in a special brochure or dictionary. The most important resolutions of TC of LAS and approved terms that are intended for a wide use in the society are published in the newspaper "Latvijas Vēstnesis" and in a central newspaper.

TC of LAS examines, adjusts, analyses and approves 400 or 500 terms per month. Some of them may be created on the spot, some demand serious consideration.

All terminology subcommissions use TRADOS MultiTerm terminology management system. They maintain the database by adding new terms and their explanations. If the term is presented in different branches, the consolidation of different meanings of it is explored and the corresponding Latvian term (terms) is proposed. A database uniting these specialized databases is created. This database is available to Internet users.

1.3 Classification systems

This section describes the use of classification systems in terminology management. Classification systems are used to organize the terms of a term collection which implies that

- a classification system should cover all 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 the concept related with a 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.

1.3.1 Lenoch universal classification system

The Lenoch classification system is a complex and fine-grained classification system covering a wide range of subject fields and is widely used for both terminology and documentation:

- http://www2.uibk.ac.at/translation/termlogy/lenoch.html

It has been developed by Dr. Lenoch (European Commission) as a subject field classification for Eurodicautom.

At the top level Lenoch works with 48 subject fields each labeled with a two-letter code:

AD	Management in the public and private sector
AG	Agriculture, fisheries, forestry – food processing industries
AR	Art
AS	Insurance
AT	Nuclear industry (with applied atomic and nuclear physics)

AU	Automation (includes telecommunications and computers)
BA	Building industry
BZ	Botany and zoology
CE	The European Communities
CH	Chemistry
CO	Commerce – movement of goods
DE	Defense
DI	Documentation and information
DO	Domestic economy
EC	Economics
ED	Education
EL	Electrical engineering and energy
EN	Environment
ER	Earth resources – energy
FI	Financial affairs – taxation – customs
GE	Generic civilization – heritage
GO	The cosmos
HI	History, ethnology, manners and customs
IC	The chemical industry
IN	Various industries and crafts
JU	Law
LA	Language and literature
MA	Mathematics
ME	Medicine
MG	Mechanical engineering
MI	Mining
NO	Standards, measures and testing
OO	News-systems and communications
OR	International organizations
PG	Printing and publishing
PH	Physics
PO	Politics
RP	Religion and philosophy
RS	Risk Management – security
SC	Co-operatives
SI	Iron and steel industries
SO	Man and society
SP	Sports, entertainments and leisure
ST	Statistics
TE	Technical and industry in general
TR	Transport
TS	Land and property
TV	Labour

Each of the subject fields has a number of subclasses from about 10 to about 100, which are labeled with a letter or digit following the subject field code.

As a direct consequence of its fine-grained structure, some term bank administrators have found Lenoch too complex to manage, making it hard for the translator to classify the terms correctly. Still, Lenoch is one of the most widely used classification systems and constitutes the basic classification, possibly with some adjustments, of many term banks.

1.3.2 Eurovoc

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. Eurovoc Thesaurus can be used also as a terminology classification system.

Eurovoc exists in 16 official languages of the European Union (Spanish, Czech, Danish, German, Greek, English, French, Italian, Lithuanian, Hungarian, Dutch, Portuguese, Slovak, Slovene, Finnish and Swedish). In addition to these versions, it has been translated by the parliaments of a number of countries (Albania, Croatia, Latvia, Poland, Romania and Russia).

At the top level Eurovoc deals with the following 21 subject fields which are all of importance for the activities of the European institutions:

04	POLITICS
08	INTERNATIONAL RELATIONS
10	EUROPEAN COMMUNITIES
12	LAW
16	ECONOMICS
20	TRADE
24	FINANCE
28	SOCIAL QUESTIONS
32	EDUCATION AND COMMUNICATIONS
36	SCIENCE
40	BUSINESS AND COMPETITION
44	EMPLOYMENT AND WORKING CONDITIONS
48	TRANSPORT
52	ENVIRONMENT
56	AGRICULTURE, FORESTRY AND FISHERIES
60	AGRI-FOODSTUFFS
64	PRODUCTION, TECHNOLOGY AND RESEARCH
66	ENERGY
68	INDUSTRY
72	GEOGRAPHY
76	INTERNATIONAL ORGANIZATIONS

Some subject fields are more developed than others, depending on how closely related they are to the Community's focus of interest.

1.3.3 Comparison and use of terminology classification systems

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

Important features of Eurovoc are that it is a controlled vocabulary with official translations in at least 16 languages. It is also more up-to-date and is better maintained.

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

1.3.4 Local scenario

At the local level, terms from usually just one or a few closely related subject fields are prepared. It is often not of crucial importance which classification system is used. However, taking into account that terms will later be used in a national context, it is advisable to make precise distinctions between different domains also at this level.

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

1.3.5 National scenario

In the national scenario, it is important to use an advanced terminology classification system. It helps to distinguish between similar terms and similar concepts and establish whether a specific 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.

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

1.3.6 International scenario

At the 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).

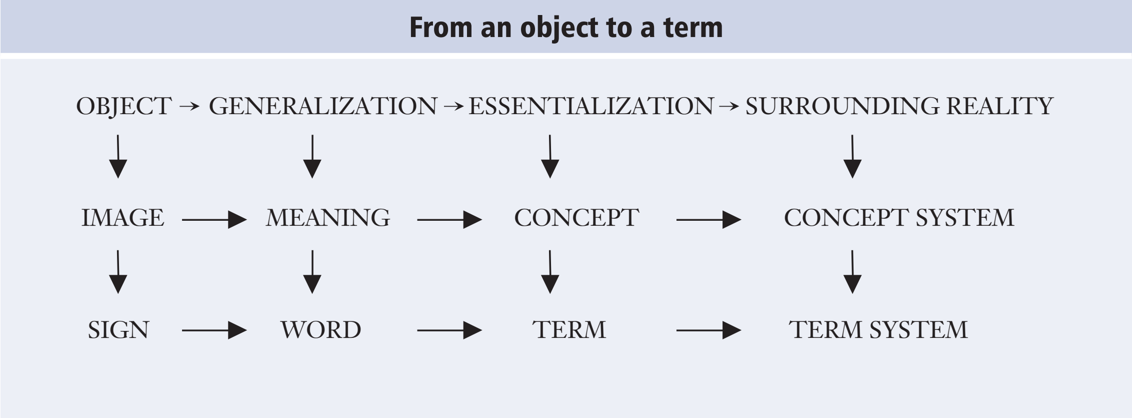
1.4 Concept analysis

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

In the term extraction process, a corpus or another collection of texts is systematically scanned for terms, their typical linguistic contexts and usages. This process can be regarded as the linguistic dimension of terminology work.

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 a 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. In the following diagram the arrows show the directions of the processes:



From an object to a term

OBJECT → GENERALIZATION → ESSENTIALIZATION → SURROUNDING REALITY

IMAGE → MEANING → CONCEPT → CONCEPT SYSTEM

SIGN → WORD → TERM → TERM SYSTEM

Figure 1: The process of concept analysis.

Concept 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 employed in 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.

Concept analysis is necessary in order 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 not valid (not preferred) variants, etc.)
- term equivalents across two, three or more languages, etc.
- Concept analysis of the extracted terms is generally carried out in two basic working situations:
- systematic compilation of the subject field terminology supplemented by conceptual analysis of terms, term groups and resulting in subject field term systems represented in term dictionaries, larger 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.

Both above-mentioned working situations may apply at the same time. An illustrative example is the development of science and economy being a continuous process where new concepts appear and need to be named with new terms. Here, 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. On the other hand, the elaborated term and concept system (of the field in question) facilitates the finding of a new systemic term for a new concept.

Creation of a concept system includes the following stages, which are to a certain extent relevant in all scenarios (viz. local, 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
- analysis of terms (from the linguistic point of view) in different sources
- preliminary assessment of the developed concept system (the assessment of the term and concept system with respect to synonymy, homonymy and polysemy relationships)
- final assessment of the developed concept system.

1.4.1 Local scenario

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.

The scope and conditions of terminology work at this level may vary between several options, for example:

- 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 within the scope of appropriate domains
- from no or variable staff unit(s) to a permanent staff group on terminology issues
- from bilingual to multilingual tasks.

The fact that organizations in a local scenario are usually only concerned with one particular domain also means the following:

- limited necessity for domain concept analysis
- no regulation by outside terminology institutions

- 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.

For best practice in the area of concept analysis, the following recommendations can be given:

- for company/project internal use at least a bilingual term database should be created and maintained
- at least a group of related terms and concepts should be analyzed
- detailed concept analysis should be performed in order to solve problems with closely related concepts that are expressed by homonyms, synonyms or polysemantic terms. (At the local level, such sets of concepts should be clarified only when they are directly related to the terms in the documents currently dealt with).

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

Dealing with translations, the source language texts constitute the basis 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.

1.4.2 National scenario

The optimal conditions of best practice in terminology work at the national level are:

- compulsory status of the official terminology stated in national legislation, e.g. the Official Language Law;
- 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 the 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.
- 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.

Unification of terms and concepts in the frame of a national language has to be done before the unification of terms and concepts on the international scale.

The following working conditions and actions can be recommended (not exhaustively):

- involvement of a wide range of subject field experts (in order to ensure a broad domain coverage), a sufficient number of high level linguists and skilled terminologists
- close collaboration among these above-mentioned professionals

- implementation of scientific-based requirements for terms and term and concept systems including harmonization on three basic levels, i.e. intra-domain, inter-domain and national
- development of a unified, regularly updated national term database for users in a national framework and for outside users
- close collaboration with terminology research institutes
- introduction of compulsory terminology courses in professional study programs
- adequate financing.

1.4.3 International scenario

International level primarily in this context means that the appropriate principles accepted for international coordination of terminology work must be observed. 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.

Therefore, the primary goals particularly relevant at this level are as follows:

- develop high quality term and concept systems in each related partner language
- provide unambiguous and consistent definitions.

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

The recommendations relevant to this level are the following:

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

1.5 Data structure and data categories

Irrespective of the terminology scenario, the principal rule recommended is to observe the basic data modelling principles as described in ISO standard 12200:1999 and 12620:1999. This will ensure exchangeability and facilitate recognition and comprehension of data categories for new or outside users. Principles of these ISO standards require that the term entries:

- are concept oriented
- contain a rather broad selection of data categories that permits the necessary level of detail (data categories and the contents of these should reflect each other precisely)
- permit full descriptions of each term

1.5.1 Local scenario

In a local scenario, the typical conditions and goals that are significant for the design of a data structure are: tight time frames, orientation towards translation, exchangeability as a high priority and restriction to a few domains. These criteria speak in favour of a highly customized and only moderately exhaustive data structure where data categories are consistent with the requirements of the particular application area and have a translation related focus.

The focus on translation requirements implies coverage of more than one language. It must therefore be considered whether descriptive concept-related information (definition or explanation) is necessary for each language or only for one language. If the term collection is multilingual, a definition for each language is usually necessary. If the term collection is only bilingual, it may not be necessary.

A focus on translation requirements also indicates inclusion of data categories permitting sufficient information about the use of a term, for example different types of grammar information, context information and collocation information. Some translation settings may also require grammar information for each word of a term. Furthermore, it is often considered very important to document the degree of equivalence between terms of different languages. Data categories that could be relevant in this respect are, for example, false friend, directionality and transfer comment.

The below data structure containing 3 levels reflects a multilingual terminology setting permitting, for example, concept descriptive information for each language but grammar information only for a term as a whole. In multilingual and perhaps also bilingual terminology settings, insertion of a word level permitting grammar information for each word of a term could be considered. In some bilingual terminology settings, having concept related information for only one language could be considered. Consequently, the data structure in a bilingual framework may include only 2 levels, namely, concept and term levels.

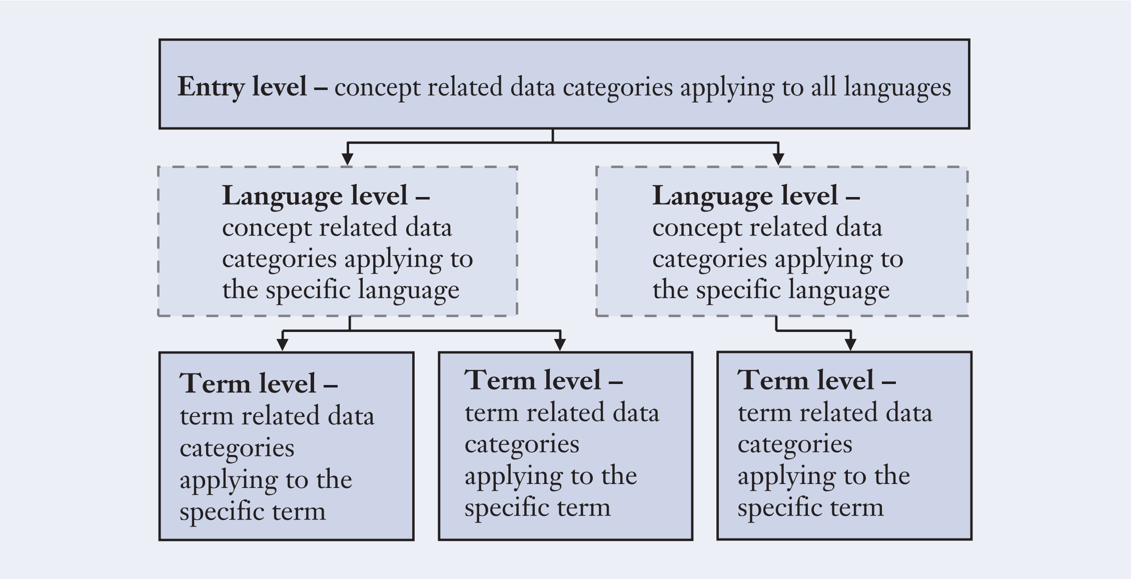

Entry level – concept related data categories applying to all languages
Language level – concept related data categories applying to the specific language
Language level – concept related data categories applying to the specific language
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term

Figure 2: Example of a data structure in a multilingual terminology setting.

1.5.2 National scenario

Conditions and goals influencing the design of a data structure in the national scenario are adequate financial support, exchangeability, broad domain coverage and high quality in general terms. Besides, a national term collection is aimed at terminology coordination and regulation rather than at translation. These criteria point towards a data structure that permits an exhaustive selection of data categories covering very different user requirements and enabling users to develop entries for very different purposes and of a very high quality.

This implies that the data structure should typically contain 2 levels: concept and term levels (at least when the term collection is monolingual) and that data categories should represent a wide selection of information types and include term status qualifiers reflecting for example acceptability, approval or applicability of a term in a given context. An example of a term status qualifier is normative authorization which is assigned by an authoritative body and includes qualifiers such as standardized term, preferred term, admitted term and deprecated term.

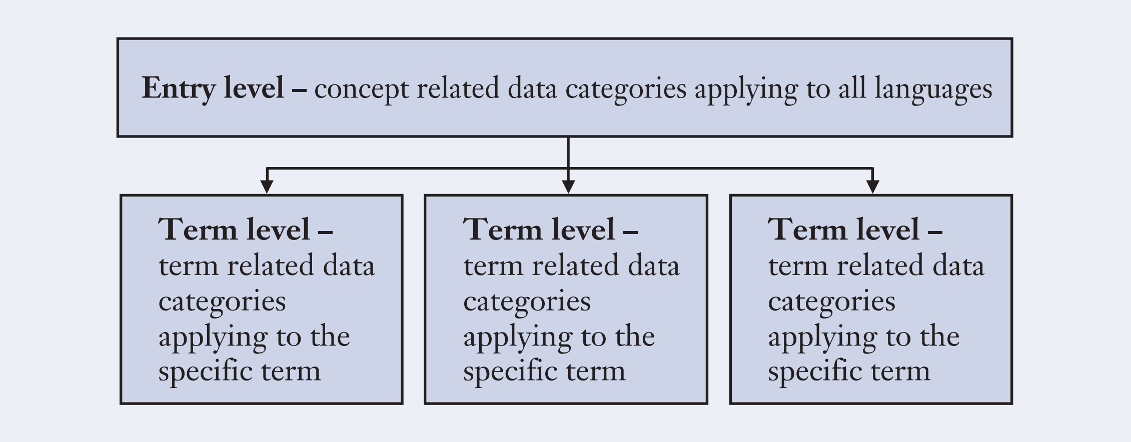

Entry level – concept related data categories applying to all languages
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term

Figure 3: Example of a data structure containing two levels.

1.5.3 International scenario

A crucial difference is of the international scenario is that international terminology cooperation is multilingual by nature. Otherwise, the criteria considered important in an international scenario are very similar to those considered important in a national scenario. Therefore it is recommended that the data structure should include four levels permitting concept descriptive information for each language, translation related information types and grammar information for each word of a term, as shown in the following illustration.

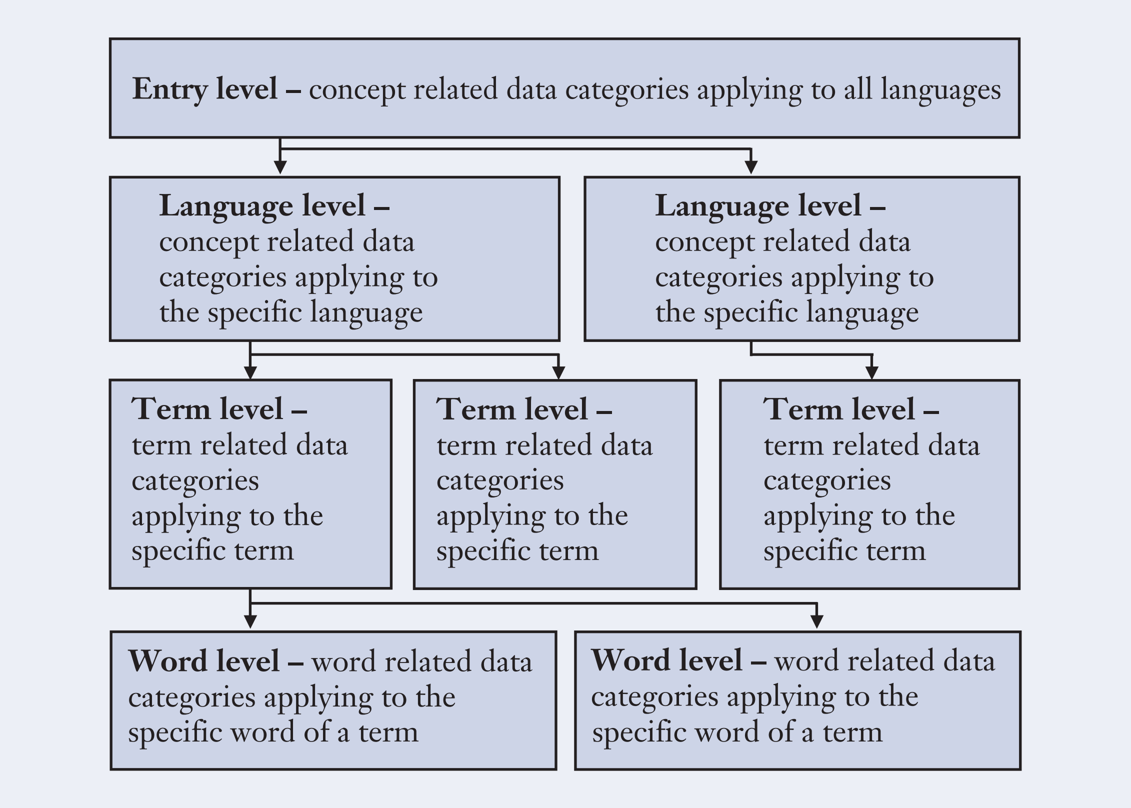

Entry level – concept related data categories applying to all languages
Language level – concept related data categories applying to the specific language
Language level – concept related data categories applying to the specific language
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term
Term level – term related data categories applying to the specific term
Word level – word related data categories applying to the specific word of a term
Word level – word related data categories applying to the specific word of a term

Figure 4: Example of a data structure in an international scenario.

1.5.4 IATE data categories: a case study

IATE (Inter-Active Terminology for Europe) is an online terminology database for all official institutions and agencies within the European Union. The following table lists IATE data categories in terminological entries:

Levels	IATE data fields	
Language independent level	LIL_RECORD INSTITUTION AUTHOR PROPOSER MARKED_FOR_DELETION_MERGING CONFIDENTIALITY DATE_MADE_CONF MADE_CONF_BY_USER	CREATION_DATE CHANGED_BY CHANGE_DATE CHANGED_IN_FIELDS DOMAIN DOMAIN_NOTE ORIGIN ORIGIN_NOTE PROBLEM_LANG_CODE COLLECTION CROSS_REFERENCE GRAPHICS
Language level	LIL_RECORD AUTHOR TERM TERM_TYPE LOOKUP_FORM OBSOLETE TL_COMMENT COMMENT_CONF DATE_COMMENT_MADE_CONF COMMENT_MADE_CONF_BY_USER RELIABILITY_VALUE TERM_REF TERM_REF_CONF	LANGUAGE_USAGE LANG_USAGE_REF LANGUSE_REF_CONF REGIONAL_USAGE REG_USAGE_REF REGUSE_REF_CONF CONTEXT CONTEXT_REF CONTEXT_REF_CONF GENDER PART_OF_SPEECH
Term level (includes word level information)	TL_RECORD AUTHOR PROPOSER INSTITUTION CREATION_DATE CHANGED_BY CHANGE_DATE	CHANGED_IN_FIELDS MARKED_FOR_DELETION_MERGING INITIAL_SOURCE VALIDATION_STATUS STAGE CYCLE

1.5.5 PolTerm data categories: a case study

PolTerm is the terminology bank of legal terminology in Polish-English and Polish-German. Due to increasing harmonization of the Polish law with the European Union laws and the data structure recommendations drawn up within the framework of the ETB project, it was decided to adopt the following data structure:

Types of data categories	Data categories
House-keeping data categories	Creation Date Change Date Inputter Subset owner Entry Number
Linguistic data categories	Term PL Term EN Language symbol Alternative equivalents (acceptable English equivalents other than the one in „Term EN" field)
Conceptual data categories	Subject field Reference (to a term: a specific-subject Act of Parliament) Definition PL Definition EN Explanation (translation of a definition of a Polish term (Term PL) into English: either to show the system-bound specificity of a particular Polish legal concept or to supplement the original English-source definition of a particular English equivalent (Term EN)
Miscellaneous	Reference (apart from Reference to Term PL and Term EN fields, also to: Definition PL/EN, Explanation and Alternative equivalents fields)

1.6 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 vary considerably, which means that the structure of different terminology databases exhibits 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 the costly individual programming of conversion routines, ISO/TC 37 has developed three international standards related to terminology interchange. ISO 12200 and ISO 12620 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 is related to the terminology markup framework TMF enabling to specify interoperable markup languages on the basis of 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 (TermBase eXchange), a 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 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.

1.6.1 Local scenario

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 procedures are much more complex. A customized specific exchange routine between two termbases can be programmed, but the more terminology resources are involved the larger number of additional exchange routines are necessary.

Although such a customized format can solve all needs for terminology interchange with a specific organization, it is strongly recommended to apply standardized exchange formats like TBX even in this exchange scenario, since sooner or later new termbases (or other applications) may come along and the need for terminology interchange with other organizations may arise.

1.6.2 National scenario

In a national scenario, 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.

1.6.3 International scenario

The recommendations for terminology interchange in an international scenario are identical to those in a national scenario. The multilingual aspect of terminology resources and the involvement of perhaps several partners with different terminology management systems, different data structures etc. add to the importance of applying a standardized format as TBX.

1.7 Validation workflow

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

The validation of term entries consists of two main steps: the formal structure on the one hand and 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 qualifications of the people involved. Verification of terminology in many languages carried out by experts from different countries requires a huge coordination effort.

1.7.1 Local scenario

In a local scenario, a restricted budget and a tight timeframe are more likely than in national or international scenarios. The ranking and requirements for terminology work can differ widely as they are also linked to the nature of the organization's core business. For the purpose of a schematic representation it is assumed that, in a local scenario, terminology work is bi- or multilingual and covers a very limited number of subject fields. It is performed by a small number of in-house employees without special terminology management tools. Beside quality, the speed of the terminology work and up-to-dateness of the data are just as important.

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 – 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 of every term

In case of pure word lists or databases defined in a rather simple way, 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 a far more complicated step contains the content check for every entry. It should 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.

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 can often be recommended to organizations of all scenarios, even if the data is for internal use only.

In a local scenario, the validation routine is in general carried out by internal employees. The maintenance of the terminology database should be assigned to a translator, if no terminologist is involved. In an ideal case the creator of an entry should not control his or her own work, although this might be inevitable due to a lack of qualified staff. The formal check should be performed by a translator without knowledge of the subject field, if possible assisted by technical means like automatic spell checking, automatic date format control etc.

The content validation should 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 should 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 finalized.

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 pre-defined feedback form would be helpful.

1.7.2 National scenario

In a national scenario, mono- or bilingual terminology is created for at least national dissemination. Many language and subject field experts from different institutions may take part in the terminology creation and validation workflow, which requires a high organizational effort. In this scenario, high quality has the highest priority, time and financial restrictions being certainly less severe than they may be in a local scenario.

Terminology tools are available — an electronic terminology management system, internet/intranet access, and possibly an integrated or external project management tool. The terminology work is usually based on sound research principles, and international standards are taken into account. So a well-considered and sophisticated database design based on ISO 12620 is taken for granted.

It is recommended that 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.

It is recommended that the content check of every entry comprises 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.

For coordination purposes the competencies of all persons involved in the validation workflow have to be well defined. A terminologist should be appointed the project leader responsible for all validation processes. The creator of an entry must not check, but he or she may correct the data according to the feedback of the reviewing persons.

The control of the formal requirements should be performed by a translator or terminologist without knowledge of the subject field, as far as it cannot be run automatically.

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

Feedback mechanisms

Information in the validation process should be processed electronically to the extent that this is possible. The project management tool might provide for a feedback routine, facilitating the coordination of the validation procedure. The entry mask for the terminological data should 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 should 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 clears errors on the spot.

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 should be drafted and distributed to the experts. 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.

The terminological entry is validated 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.

1.7.3 International scenario

In an international scenario, 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 than in a national framework, and high quality of the terminology has absolute priority. Nevertheless, the validation workflow is similar to the validation workflow in a national scenario.

A typical validation workflow on national or international level is provided in this illustration:

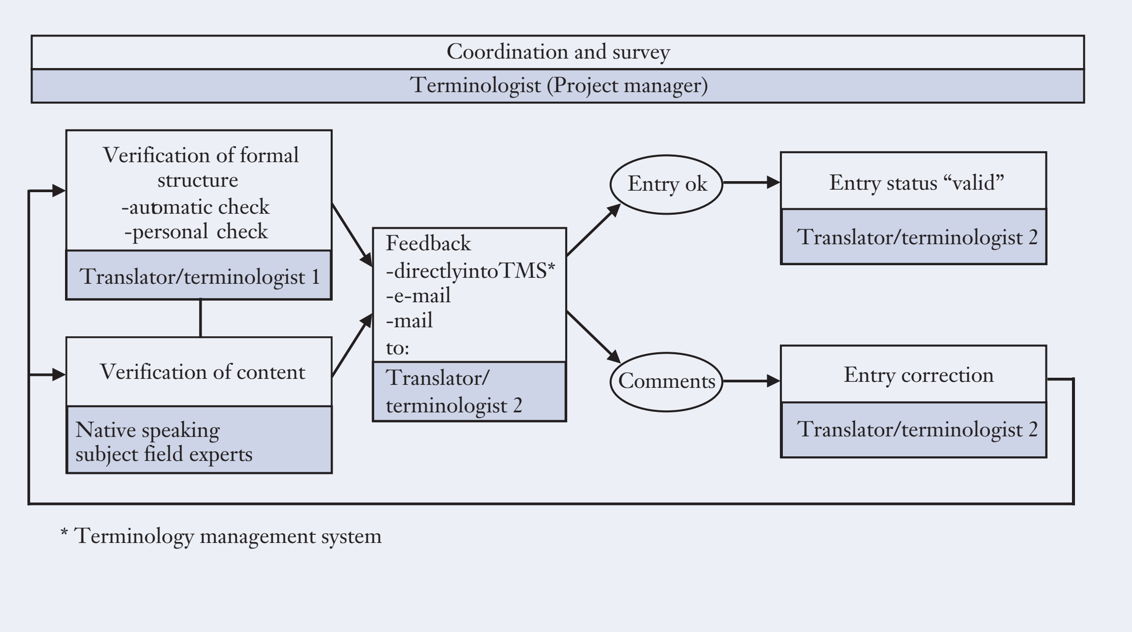

Coordination and survey
Terminologist (Project manager)
Verification of formal structure
-automatic check
-personal check
Translator/terminologist 1
Verification of content
Native speaking subject field experts
Feedback
-directlyintoTMS*
-e-mail
-mail
to:
Translator/ terminologist 2
Entry ok
Comments
Entry status "valid"
Translator/terminologist 2
Entry correction
Translator/terminologist 2
* Terminology management system

Figure 5: A typical validation workflow on national or international level.

As already mentioned, validation concerns the final check of formal structure and content of the entries. In order to support, facilitate and in some ways provide the schematic framework of validation, it is also important to define validity criteria and writing rules.

1.7.4 Validity of term entries in IATE: a case study

A terminological entry must meet certain criteria with relation to credibility, relevance and minimum amount of information.

The IATE database should only contain relevant entries. A relevant entry concerns a drafting, translational or interpreting problem in an area of relevance for the Community. A relevant entry should not be an everyday term since the inclusion of such might lead to problems with duplication of information and the definition of the boundaries of the entries. Only everyday terms that have an added value in Community documents compared to their definitions in language dictionaries should be entered.

Entries in IATE are by definition valid and correct unless they are marked 'deprecated'.

A term should be as concise as possible, i.e. it should be the smallest unit that can designate a given concept exactly. Complex terms or expressions can often be broken down into various concepts and lack of conciseness should be avoided.

1.7.5 Validity of term entries in PolTerm: a case study

For an entry to be valid, the following mandatory terminological data categories apply in translation-oriented terminography (ISO/DIS 12 616.2: 8): term, language symbol, source. In the case of the PolTerm terminological database, these categories translate into: <Term EN>/<Term PL> fields' content; <Term EN>/<Term PL> fields' names; <Reference> field's content as regards a Polish term in relation to its Act-specific origin within the PolTerm Translation Memory; <Reference> field's content: a crucial data category in connection with English equivalents of the given Polish source terms.

1.7.6 Validity of term entries in Estonian Legal Language Centre: a case study

The minimum set of data categories that form a valid entry (in addition to categories that are necessary for the formal consistency of a database) is the following:

- Term in the source language
- Subject field
- Definition in the source language
- Context of the source language

1.7.7 IATE writing rules: a case study

The writing rules of IATE cover all fields of the data creation interface except fields that are filled in automatically or that are filled in by choosing a single item from a pick-list.

Levels	Data categories	Writing rules
Entry level	Domains and domain note	An entry should only belong to one or few domains. Several domains in one entry may indicate that the entry covers more than one concept. It should also be checked when adding new data to an existing entry whether the context and domain are the same as the ones already there. Narrower descriptors of domains can be added in domain note.
	Origin and origin note	A country name can be selected here to indicate the geographical origin of a concept, if necessary supplied in the origin note by a more specific indication of the political, cultural, ethnic or religious origin.
	Problem language	The problem language is the language of the term in which it was created and serves as the basis for addition of other languages.
	Proposed by	Name of person who proposed the entry if this is not the creator of the entry.
	Cross-references	Cross-references are links to related concepts (note that synonyms should be included in the entry itself). The type of relation that holds between the concepts should be indicated (selected from a list).
	Collections	The names of term collections should include an institution ID and a short description of the collection.
	Graphics	Relevant graphic files can be included with a short description.
Language level	Definition and note	A definition should follow the principle of substitution and should be broad enough to identify a concept in a general context. Further explanation that is not part of the definition must be placed in the note.
	Related material	A list of relevant material apart from the main references.
	Graphics	Relevant graphic files can be included with a short description.
Term level	Term	The term should be concise, i.e. should consist of the smallest indivisible part that designates the concept. Complex expressions should be separated into their constituent parts and an entry should be created for each of the concepts involved. All terms should be correct, not recommended terms should be marked as ‘deprecated’ followed by an explanation in the note field. Specific rules for how to write for each part of speech and more general rules that hold for all word classes. The field for grammatical information can be used for exception or specific indications. In the case of the lack of a definitive term (or title etc.) a provisional solution should be proposed by the terminologist and explained in a note. The term should be updated as soon as possible.
	Short form	The short form of a name or title should be included where appropriate.

Levels	Data categories	Writing rules
	Phrase	"Phrases" which are frequently found in Community documents, have a standard translation and pose a problem for translation should be included.
	Abbreviation	Should be written according to the language specific rules.
	Formula	Chemical formulae, mathematical and other scientific expressions should be written according to international standards.
	Term Group	Term group numbers indicate synonyms that are morphologically related.
	Context	The purpose of this is to demonstrate how a term is used in context.
	Language usage	Provides information on usage and style or level of language.
	Regional usage	Indicates whether a term reflects regional usage.
	Customers	Customer names should ideally include an 'institution ID'.
	Lookup forms	Forms of the terms like spelling variations or inflected forms that should be made searchable.
	Proposed by	User name of the proposer of the term level information if different from the creator.
	References	See Source Identification.

CHAPTER 2

TERMINOLOGY INFRASTRUCTURE AND STANDARDIZATION

This chapter provides insights into the state of terminology work in selected new EU member countries, briefly characterizes the main players in the field of terminology in general, as well as summarizes on major terminology tools and standards. The material presented in this chapter is relevant background information gathered and researched as part of the EuroTermBase project.

2.1 State of terminology management in selected new EU member countries

In the new EU member countries, much like the old ones, stakeholders in terminology development are public institutions, universities, technical societies as well as representatives of the private sector. This chapter lists major institutions in selected countries: Estonia, Hungary, Latvia, Lithuania, and Poland.

Overall, terminology situation in these countries is characterized by terminology resource fragmentation across different institutions, inconsistency and lack of coordination in terminology development, as well as structural and technical incompatibility. A lot of terminology data is available only in the form of printed dictionaries and bulletins or stored in card files. The transformation from centralized terminology development during Soviet time with the focus on Russian language to requirements of market economies is still not fully completed. At the same time, positive trends do exist. Grass-root terminology development activities are carried out by field experts, ensuring that new terminology is created. There is usually at least one or several legal or governmental bodies involved in terminology management. Accession to EU has served as a major impetus for terminology work.

2.1.1 Estonia

There is no central body responsible for the terminology development in Estonia. Terminology is developed by specialists of the respective field. Typically, a field has a voluntary terminology committee that harmonizes the terminology of that field. The committee is not a legal entity; it may operate under a ministry, a non-profit organization, a university, etc.

The following legal entities are responsible for some aspects of terminology work:

Organizations	Web addresses	Descriptions
Estonian Terminology Association (ETER)	www.eter.ee	A non-profit organization, with over 60 individual and collective members. The mission of ETER is to coordinate the work of LSP and terminology in Estonia and develop term collections.
Estonian Legal Language Centre	www.legaltext.ee	A public organization under the governance of the Ministry of Justice. The mission of the ELLC is to meet the legislative translation and terminology development needs of the Estonian Government. The ELLC fulfils its mission by the following main activities: • translation of Estonian legislation into English; • translation of EC legislation into Estonian; • creation, administration, dissemination of a full-text database of legal translations and of a terminology database.
Institute of the Estonian Language	www.eki.ee	A public research and development organization under the Ministry of Education and Science. The mission of EKI is to research Estonian (modern Estonian, dialects, history of language, LSP etc), including terminology in Estonian. EKI is also responsible for language planning.

A number of major Estonian terminology resources are shown in the following table:

Names/addresses	Descriptions
keeleveeb.edu.ee	General access portal, containing dictionaries and links to various Estonian mono- and bilingual dictionaries, both general-purpose and specialized.
www.keelevara.ee	Linguistic portal requiring registration and payment for using some of its dictionaries.
www.legaltext.ee	Esterm database
www.eoy.ee	Names of the birds of the World
www.loodus.ee/eurolinnud	List of Birds of Europe
www.ut.ee/taimenimed	Estonian Plant Names
www.pangaliit.ee/pangandusinfo/sonastik	Lexicon of bank terminology
www.matk.ee/termin/sonastik.htm	Lexicon of ramblers´ terminology

2.1.2 Hungary

In Hungary, the first major government effort in the terminology field was started by the Ministry of Justice in 1997, when the Translation Coordination Unit was established. The aim was the creation of the official Hungarian terminology database of the European Union. Following the EU accession in 2004, Ministry of Justice decided to upkeep its TCU, but the amount of work and staff has been on the decrease. Terminology work was being done in several organizations, yet there was hardly any cooperation between the actors. Several organizations initiated a broad terminology dialogue on a national level. Finally, with the establishment of the Terminology Council of the Hungarian Language (MATT) in May 2005, all terminology work is brought to a national, standardized level, integrating all previous efforts.

Organizations	Descriptions
Translation Coordination Unit (TCU)	Established in 1997, under Ministry of Justice, High Department for European Community Law. Translation was procured from translation companies. TCU consisted of 4-9 terminologists and lawyer-linguists. Created a terminology database (Termin) of 23,000 entries, available on Internet.
Terminology Council of the Hungarian Language (MATT)	Established in 2005, MATT performs research, education and training, on terminology and language policy, drafts strategic recommendations, coordinates terminology work nationally, cooperates with international terminology organizations.

2.1.3 Latvia

In Latvia, the main institution for the development of unified, coordinated and harmonized multi-branch terminology since 1919 is the Terminology Commission of the Latvian Academy of Sciences (TC of LAS). Decisions taken by the TC of LAS have the status of normative documents, and terms approved by TC of LAS are official. Besides the Terminology Commission, there are a multitude of other organizations involved in some aspects of terminology work:

Organizations	Descriptions
Terminology Commission of the Latvian Academy of Sciences (TC of LAS)	As per State Language Law, TC of LAS is responsible for development of a uniform national terminology system and coordination of terminology development. At present it consists of 26 subject-field terminology subcommissions.
The State Language Commission (SLC)	Established under the auspices of the President of Latvia, serves as the main institution determining the state language policy in Latvia.
The State Language Agency (SLA)	A government body under the Ministry of Education and Science. Among its major tasks are consulting and promotion of Latvian as the state language.
The State Language Centre (SLC)	An institution under the Minister of Justice. The purpose of the Centre is implementation of the state policy, performing supervision over the observance of regulative acts and control in the field of the state language use.

The Latvian Language Institute (LLI)	A research institute under the authority of the Latvian University (LU). A structural unit of the LLI is a Terminology Department.
The Translation and Terminology Centre (TTC)	Founded in 1997 by the Ministry of Foreign Affairs, with the main function of translating EU legislative acts. Since 2005, TTC is under the Ministry of Education and Science. The aim of TTC is providing translations of documents of state and international organizations for the purposes of state administration and the society, as well as to submit proposals for development and standardization of terminology.
Latvian Standardization Organization „Latvijas Standarts” (LVS)	Founded in 1999, is the national standardization body, and its main tasks are to provide information on standardization, develop the national standards, adapt international standards and maintain the register of adapted Latvian standards.
Tilde	Established in 1991, is a leading Baltic IT company specializing in language technologies, multilingual and Internet software, localization. As a member of the Information Technology and Telecommunications Terminology Subcommittee of the TC of LAS, Tilde actively participates in terminology development process.

The major Latvian terminology databases are described in the following table:

Names/addresses	Descriptions
Termnet, www.termnet.lv	Hosted by Tilde in cooperation with Academy of Sciences of Latvia, provides access to about 145 000 terms in different domains. It is also a portal where all new official terms get posted, and users can post comments. Many entries contain terms in up to 4 languages and a definition.
TTC database, completeddb.ttc.lv	Hosted by the Translation and Terminology Centre. It was started by digitalizing terms created by Terminology Commission of Latvian Academy of Sciences (TC of LAS) in last 10 years as well as adding terms created in EU and NATO materials translation process. Many entries contain terms in up to 4 languages and/or usage examples (context). Database is continously updated with in-house created terms and several thousand terms approved by TC of LAS.
www.termini.lv	Hosted by TehnoMedia, provides IT&T and physics terms. Mostly interesting as an advanced term search engine capable to recognize similar words, misspelled words and words derived from the same root.

2.1.4 Lithuania

Terminology work in Lithuania is centralized; there are three main institutions that work in this field: the Centre of Terminology at the Institute of Lithuanian Language, The State Commission of the Lithuanian Language and The Standards Board. In addition, specialists of various fields carry out terminology work at Lithuanian universities.

Organizations	Descriptions
The Centre of Terminology at the Institute of the Lithuanian Language	Terminological work at the Institute has been carried out from the establishment of the Institute in 1941. It has the following main tasks: research, establishment of terminology principles and norms, terminology development, training of terminologists, consulting. Until 1996, responsible for approval of terminology dictionaries and standards. The Centre publishes an annual terminological magazine called Terminologija (since 1994).
The State Commission of the Lithuanian Language (SCLL)	Established in 1990, decides on language policy, normalization and standardization of the Lithuanian language and implementation of the State language law. Is the creator and manager of the national term bank (the Term Bank of the Republic of Lithuania). The Terminology Sub-committee of SCLL reviews and approves terminological dictionaries, terminological standards, collections of terms, terms for the Term Bank of the Republic of Lithuania, etc.
The Lithuanian Standards Board (Technical committee 37 Terminology)	Established in 1990, the national standards body. Consists of 6 sub-committees which are responsible for the expertise of Lithuanian terms in the projects of standards. The first sub-committee also carries out the adoption of standards prepared by ISO/TC 37 and is responsible for participation in ISO/TC 37 activity.

The major Lithuanian terminology databases are described in the following table:

Names/addresses	Descriptions
Lithuanian Language Term Base (Lietuvių kalbos terminynas), www.terminynas.lt	Created on the basis of dictionaries of terms. The Term Base is a source of information for linguists, translators and editors. It is accessible on Internet for registered users.
Term Bank of the Republic of Lithuania, http://terminai.vlkk.lt:10001/pls/tb/tb.search	The State Commission of the Lithuanian Language together with the Chancellery of the Parliament took the initiative to create a State Bank of Terms; the law on the Term Bank of the Republic of Lithuania was passed in 2003. The purpose of the Term Bank is to ensure a consistent usage of normalized Lithuanian terms, especially in the legislative documents of the Republic of Lithuania, to create a common informational system for various state institutions with the possibility for other persons and legal entities to get connected to it and to provide data to it.

2.1.5 Poland

There are a number of institutions in Poland dealing either with the methodology of terminology, creation of terminology collections in particular fields, or maintenance of terminology resources.

Organizations	Descriptions
Polish Standardization Committee 'PKN' (Polski Komitet Normalizacyjny)	Comprehensive terminology work in all fields of terminology work: methodology, creating and maintaining terminology resources.
The Office of the European Integration Committee 'UKIE' (Urząd Komitetu Integracji Europejskiej)	Established in 1996, Department for Translation created a bank of legal terms and phrases used in Polish translation of EU legal instruments.
Council for the Polish Language	Established in 1999 as a committee of the Polish Academy of Sciences dealing mainly with the methodology of creating specialized terminology.
International Specialized Terminology Organization 'ISTO'	Founded in 2004, focusing on international terminology work, a holder of a certain number of specialized terminology dictionaries.
Polish Society of Sworn and Specialized Translators 'TEPIS', in cooperation with the 'Translegis Publishing House'	A term bank named 'PolTerm' is maintained by the Translegis Publishing House.

The list of major terminology resources for Polish is provided in this table:

Names of termbanks	Descriptions
Unified Terminology Bank 'BTZ', owned by PKN	About 77,000 records concerning terms extracted from Polish standards. The 'BTZ' Bank is maintained in the ISIS program and includes the following information in their records: term, definition, remarks, synonyms, foreign language equivalents, etc. Recently, the PKN Committee has been involved in restructuring their data base through changing their software system into Integrated Computational System 'ZSI' which will take some time before making their data base available to the internet users.
Terminology collections published by the PKN Terminology Bank Committee	Series of three publications containing collections of terms extracted from respective standards: work safety, information technology, environmental protection.
UKIE terminology data base	The UKIE terminology data base contains terminology collected while translating EU legal instruments into the Polish language. The total number of records has reached 8000.

UKIE Terminology collections	Glossaries prepared at the European Commission constitute a part of these collections. UKIE Translation Department's published four-language glossaries:
	Glossary of the Community Customs Law Terms (CCC);
	Glossary of the Treaty on European Union Terms (TUE);
	Glossary of the European Agreement Terms (EA);
	Glossary of the Terms Relating to Internal Market (INT);
	Glossary: Economics – Finance – Money (EKOFIN);
	Glossary: Regional Policy (REG);
	Glossary: Regional Report of the European Commission (RAPORT)

PolTerm	Bilingual LSP-corpus-based terminology collection. The LSP electronic corpus contains 42 consolidated texts of Polish legislative Acts and their English and German translations. Current number of entries amount to 10,500 and cover a wide range of branches of law.

2.2 Termbanks, bodies and networks

This section provides information about several major players in the international terminology area, as well as important terminology resources especially for the languages of "old" EU countries. It looks also at a few global companies and European projects as case studies for terminology management.

2.2.1 IATE: Inter-Active Terminology for Europe

IATE is a single multilingual terminology database for the European Union. The EU Translation Centre launched the "IATE" ("**I**nter-**A**gency **T**erminology **E**xchange") project in 1999; its initial objective was to create an infrastructure for the management of terminology for the Centre and the decentralized agencies of the Union. The other translation services joined this initiative in the same year and gave the project its truly inter-institutional status. Given this change in scope, the acronym today stands for "**I**nter-**A**ctive **T**erminology for **E**urope".

The operational phase of the IATE project began in the summer of 2004. The system currently offers the following:

- One common database for all institutions and agencies containing all legacy data;
- Online access in read and write mode, i.e. the possibility for users to carry out modifications, add entries directly to the central database and thus allow their colleagues to profit from this work immediately;
- Validation procedure to ensure quality. Possibility to define validation cycles, validation stages, user profiles, user roles etc. for each participating institution and agency;
- Management tools (e.g. for user management, data consolidation):
- Features for large scale processing (export and import of data);
- Reporting and auditing tools, e.g. the possibility to trace modifications in terminological entries;

- A messaging system as the communication mechanism between the actors in the terminology workflow.

The project partners who use and jointly finance the IATE database are:

- European Commission
- Council
- Parliament
- Court of Auditors
- Economic and Social Committee
- Committee of the Regions
- Court of Justice
- Translation Centre for the Bodies of the EU
- European Investment Bank
- European Central Bank

2.2.2 Infoterm: a network of terminology centres

The International Information Centre for Terminology (Infoterm) was founded in 1971 by UNESCO, the United Nations Educational, Scientific and Cultural Organization, with the objective to support and coordinate international cooperation in the field of terminology. In 1996, Infoterm was reorganized and established as an independent non-profit organization.

Infoterm's mission is to promote and support the cooperation of existing and the establishment of new terminology centres and networks with the general aim to improve specialist communication, knowledge transfer and provision of content with a view to facilitate the participation of all in the global multilingual knowledge society.

In order to achieve this objective, Infoterm members cooperate in organizing a world-wide network of terminology centres and terminology networks with a view to:

- disseminating information on terminological activities as well as enhancing the awareness for the importance of terminology in all spheres of society,
- furthering the preparation of reusable terminologies by subject-field specialists in cooperation with terminologists,
- sharing the expertise regarding harmonized methods and guidelines for terminology management, the management of terminology centres, and for the use of terminological data, methods and tools in all applications where specialized information and knowledge are involved.

Infoterm's vision is to organize the methodological and organizational basis for a most efficient and effective preparation of terminologies in the form of net-based distributed cooperative terminology work under a comprehensive content management approach guaranteeing semantic interoperability across all application fields.

Infoterm members are either terminology organizations/institutions or specialized organizations/institutions with major terminological activities, which can be considered to be authorities in their field. They are public institutions, intergovernmental organizations and non-profit organizations. Members cooperate in organizing a worldwide network of terminology centres and terminology networks.

Infoterm publishes, besides books on terminological issues, several quarterlies – the Infoterm Newsletter (INL), BiblioTerm (BIT) informing its readership about the latest publications in the field of terminology, StandardTerm (STT) providing up-to-date information on standardization in the field of terminology, including standardized guidelines for elaborating terminologies, and Terminology Standardization and Harmonization (TSH), a joint publication of the ISO/TC 37 Secretariat and Infoterm.

2.2.3 TERMIUM: the Canadian termbank

TERMIUM, sponsored by the government of Canada, is one of the largest termbanks of the world and is maintained continuously (approx. 100,000 modifications per year). The main content features of TERMIUM are:

- 3,500,000 terms and names in English and French
- standardized English and French terminology
- 100,000 terms and names in Spanish
- information types: synonyms, acronyms, abbreviations, definitions, contexts, phraseology units, examples of usage and observations
- subject fields: "almost every field of human endeavour is covered"

The content of the database is accessible to translators, technical writers and other professionals. Several spin-off products are also developed, such as an on-line linguistic tool the TERMIUM Plus® which is built on top of the termbank, providing writing assistance facilities in English and French and giving access to 13 electronic language resources.

The experience gained in developing and maintaining TERMIUM is formulated in a comprehensive tutorial, the Pavel Terminology Tutorial which is on-line and freely accessible through the internet at the address:

- http://www.termium.gc.ca/didacticiel_tutorial/english/lesson1/index_e.html

This tutorial can be considered a documentation of best practices recognized in the development of the TERMIUM data bank.

2.2.4 TSK: the Finnish Terminology Centre

The Finnish Terminology Centre TSK (Sanastokeskus TSK) offers information and expert services related to special language terminology, vocabularies and terminology work. TSK's main activities are terminology projects, termbank activities and term service.

Since 1974, TSK has been the only national terminology centre with the responsibility of coordinating all terminological activities, taking care of the special language planning in Finland in cooperation with The Research Institute for the Languages of Finland (Kotimaisten kielten tutkimuskeskus) and producing mono- and multilingual vocabularies and terminological databases in co-operation with subject field specialists.

TSK is a member of the Nordterm association and has a representative in its steering committee. TSK has participated in the planning and organization of Nordterm's courses and seminars and in the production of the Nordterm publications that deal with terminological principles and methods used by the Nordic terminology centres and institutes.

TSK is an independent, non-profit registered association free of any financial, political or other commitments. This means that TSK offers terminological services to all customers, whether in public or private sectors, and that the needs of the customers are equally tended to. According to the statutes of the association, TSK concentrates on offering information services for public purposes and not to making a profit. The member organizations of TSK have the right to decide upon the statutory matters of the association in two annual general meetings.

As the official Finnish member body of ISO/TC 37, TSK has actively participated in the production of the international standards on terminology work.

The research and development of terminological principles and methods are among the most important activities of TSK. By teaching these principles and methods to subject field specialists and to language specialists, TSK contributes to the quality of all terminology work done in Finland. TSK organizes courses in practical terminology work and in terminology project management for all those interested in terminological issues.

TSK is a developer and user of terminological databases. TSK has several years of experience in developing methods and techniques for terminological databases and termbanks. TSK's own termbank TEPA, set up in 1985, is a multilingual database of technical terminology: in addition to Finnish terms and definitions it contains term equivalents in several languages of which English and Swedish are the most common ones. It contains now ca. 100,000 term entries, thus being one of the largest public termbanks in the world. For in-house use TSK has a termbank containing ca. 250 000 term entries.

2.2.5 Glimpses of terminology work in global companies

SAP terminology management

SAP has a one-stop terminology interface integrated into its SAP R/3 system (transaction SE63 – Translation Environment).

Similarly to other SAP products, the terminology tool also offers a wide range of user privilege management features, and terms can have different statuses of approval until they reach the status of approved.

Some SAP translators – especially the localizers – get access to this software and can suggest new terms, others are provided with bilingual MultiTerm glossaries to get help in their work. These glossaries are also available for sale for a sum of 110 euro, and contain entries in Bulgarian, Croatian, Czech, Danish, Dutch, English, Finnish, French, German, Hungarian, Italian, Norwegian, Polish, Portuguese, Russian, Slovakian, Slovenian, Spanish, Swedish and Turkish. So far, the database contains about 650,000 terminology entries in 20 European languages, and nearly 16,000 definitions of SAP concepts. SAP made public some of these definitions in English and German at

- http://help.sap.com/saphelp_glossary/en/index.htm
- http://help.sap.com/saphelp_glossary/de/index.htm

At SAP, the privileged languages are English and German, and all terms need to have equivalents in all languages.

Terminological entries are usually created by knowledge brokers and authors of texts as well as in English or German. Entries include not only software-related entries (screen captions, etc.) but also entries appearing in training course materials and marketing materials. Translators can also enter new terms, but superusers – consultants – need to approve them.

Entries include a wide range of information, including definition and part-of-speech information, but the emphasis falls on the source of the term.

At SAP, therefore, all users regardless of their nationality use the same terminology database, there are no competing databases.

Microsoft terminology management

Microsoft employs a particular term registration process. Microsoft follows a systematic approach to software-encoded terminology, which starts during development time. Developers create terminology during program design and development in an intuitive/metaphoric way. Important terms – e.g. brand names, major technology names – are also checked by other personnel, sometimes even tested in a public opinion poll. The language of all source terms is English. The creation of the initial termbase is automatic: their own localization software extracts all the string resources from the products.

Microsoft does not have real termbases, in the sense that definition is not an integral part of their terminology. They use terms to provide a consistent localization to their products.

Microsoft employs a few (1-3) terminologists for every language they provide a product version for. These terminologists create a core termbase for each and every product, building on the terminology of former products and user responses. The core termbase is then sent out to localizers, who have to create local versions of their products – and their terminology.

Most of the terms employed appear in screen captions. Termbases (in the form of source string – target string) are unique for each and every product and product version, and contain the screen captions and some help-specific terms. Non-software-related terms (which are only a few in number) are not collected in a single termbase, but Microsoft Press, the official publishing house of Microsoft, regularly updates its dictionary.

All string-termbases are available to developers for free at the following URL:

- http://www.microsoft.com/globaldev/tools/MILSGlossary.mspx

The core termbases the work is based on are not published.

2.2.6 Selected projects related to terminology field

INTERA overview

INTERA (started in January 2003, duration 2 years) was a European project with two major objectives:

- building an integrated European language resource area by connecting international, national and regional data centres,
- producing new multilingual language resources.

The first goal involves the integration of a critical mass of different types of language resources with the help of metadata descriptions and the interlinking of the resulting distributed resource repository with an existing tool repository thus enabling users to directly start suitable tools on the included resources. INTERA anticipates that this integrated and interlinked metadata description domain will facilitate the access to language resources in Europe and help professionals in industry, the eContent business, research and education, and increase the usage of the resources already available.

The second goal addresses the lack of quality of multilingual resources, especially for the less widely spoken languages, including the Balkan ones, which are of crucial importance to the development of the eContent business. INTERA goes further ahead by developing exemplary methods for their business attractive production.

Of special interest to the EuroTermBank project are the achievements in the multilingual language resource production where four parallel corpora have been created from which extraction of multilingual terms have taken place, and the procedure developed on that subject.

The corpora created are:

- Greek – English parallel corpora (4 MWs (Million words) in total, 2 MWs per language)
- Slovene – English parallel corpora (4 MWs in total, 2 MWs per language)
- Serbian – English parallel corpora (2 MWs in total, 1 MWs per language)
- Bulgarian – English parallel corpora (2 MWs in total, 1 MWs per language)

And the following domains were subject to term extraction:

- domain of law: Bulgarian, English, Greek, Serbian, Slovene
- domain of education: Bulgarian, English, Greek, Serbian
- domain of health: English, Greek, Serbian
- domain of tourism: English, Greek
- domain of environment: English, Greek.

The numbers of terminological entries are rather low, between 200 and 9000 entries per domain per language, but the procedures for corpora creation and term extraction are well described.

More information on the Intera project can be found at:

- http://www.elda.org/rubrique22.html

SALT overview

In the SALT (Standards-based Access to multilingual Lexicons and Terminologies) project a consortium of academic, government, association, and commercial groups in the United States and Europe worked together on the task of testing, refining and implementing a universal putting-together format for the interchange of terminology databases and machine translation lexicons.

The project responded to the fact that many organizations in the localization industry are using both human translation enhanced by productivity tools and machine translation (MT) with or without human post-editing. The SALT project combined two already finalized interchange formats: OLIF (Open Lexicon Interchange Format), which focuses on the interchange of data among lexbase resources from various machine translation systems, and MARTIF (ISO 12200:1999, MAchine-Readable Terminology Interchange Format), which facilitates the interchange of termbase resources with conceptual data models ranging from simple to sophisticated. The goal of SALT was to integrate lexbase and termbase resources into a new kind of database, a lex/term-base called XLT (eXchange format for Lex/Term-data). XLT is based on XML (Xtensible Markup Language), which is a data format for structured document interchange on the Web and is under development by the World Wide Web Consortium.

The SALT project was an open-source project creating open standards. Some of the results of the SALT project have been turned into ISO standards or have been integrated into revised ISO standards, and ISO IPR policies apply to these. Control of TBX has been handed over from the SALT project (by the European Commission as its legal representative) to the Localization Industry Standards Association (LISA) and its OSCAR (Open Standards for Container/Content Allowing Re-use) Special Interest Group. All work carried out by the SALT project was explicitly royalty free and all IPR donations to the SALT project were made under a royalty free license arrangement.

The overall goal of SALT was extremely practical. It was to reach the “critical mass” with XLT so that tool developers, such as Star, Trados, EP, Logos, Systran, L&H, and Xerox, would incorporate some level of XLT support in their products and various companies would provide on-going consulting services to anyone who wants to get their proprietary lex/term-data into XLT format or XLT data into their proprietary format.

Partners in the project were the Institut für Übersetzer- und Dolmetscherausbildung of the University of Vienna (Gerhard Budin) as project coordinator, the Institute for Information Management, University of Applied Sciences Cologne (IIM, Klaus-Dirk Schmitz), the Accademia Europea di Bolzano per la ricerca applicata e la formazione post-universitaria (Bruno Ciola), the University of Surrey (Khurshid Ahmad, Lee Gillam), the Laboratoire lorrain de recherche en informatique et ses applications (LORIA, Laurent Romary), the Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e. V. an der Universität des Saarlandes (Jörg Schütz) in Europe, as well as the Brigham Young University Translation Research Group (Alan K. Melby) and the Kent State University Institute for Applied Linguistics (Sue Ellen Wright) in the United States.

The project was co-funded within the EU Fifth Framework Programme during the period from January 2000 to December 2001.

ENABLER overview

The Enabler (European National Activities for Basic Language Resources) Thematic Network (EC project, was started in December 2001) aims at improving cooperation among national activities established by national authorities for providing Language Resources (LRs) for their languages. The main results within the Enabler Network were the following:

- A survey of language resources (LRs), providing a global overview of National projects and activities on LRs of all kinds. It relates to 164 different resources from various countries and languages and concerns all the facets of LRs. Both the point of view of the LR producers and of the prospective users were taken into account.
- With the aim of optimizing the process of production and sharing of (multilingual) LRs, the Network promoted the compatibility and interoperability of LRs through cooperative work with projects, committees and communities in the different fields of LRs.
- Collection of validation methodologies of LRs representing current best practice in the area.
- A description of the industrial needs of LRs, with the aim of easing the exploitation of existing LRs and collecting ideas for future LRs.
- A number of initiatives with the objective of promoting LR production and management in the years to come, improving infrastructure and coordination activities for LRs.
- The BLARK (Basic LAnguage Resource Kit) concept has been adopted and supported defining a further level The Extended Language Resource Kit (ELARK).
- The promotion of the launch of a large initiative comprising the major LRs and HLT groups in Europe and world-wide for the creation of an open and distributed infrastructure for LRs.
- Contribution to the design of an overall coordination and strategy in the field of LR. A new committee has been established in the field of Written LRs, the International Coordination Committee for Written LRs and Evaluation (ICCWLRE) continuing the Enabler mission but enlarging the scope beyond the European boundaries. This committee will cooperate with the COCOSDA (International Committee for the Coordination and Standardisation of Speech Databases and Assessment Techniques).

2.3 Terminology tools

This section provides an overview of major terminology tools used in various areas of terminology management and translation.

2.3.1 Terminology management tools

Terminology management tools are computer programmes conceived especially for the management – the recording, processing, saving, and using – of linguistic data and their application in the area of technical writing, translation and terminology work. Most of them have a database-like structure less extensive than “real” databases and allow only the comparatively easy operations and commands necessary for terminology management.

Terminology management tools can be either term-oriented, or concept-oriented, handling a special language pair, multilingual or monolingual resources. They differ in their way of data structuring and modelling. Terminology management tools are equipped with various search, filter, import and export functions. Many of the terminology management tools are integrated into a computer-aided translation environment consisting of translation memory, align tool, tag editor and, in some cases, term extracting tool, and provide interfaces to standard word processing and DTP software. Some terminology management tools are available as stand-alone solutions and as multi-user systems.

During the last years, the tendency goes towards interactive online terminology management. The new internet-based terminology management tools have a client-server architecture allowing a functional division between the server performing the proper data base functions on a central computer system, and the client responsible for the interaction with the user PC. Thus, terminologists, translators and experts scattered all over the world can work simultaneously with one terminology data base.

A number of current terminology management tools are included in the following list (please note that some of them are only available as an integrated part of a translation memory tool):

- across (Ahead Software AG, Germany)
- CATS – Computer-Aided Terminology System (CATS, Germany)
- GFT DataTerm (GFT GmbH, Germany)
- Lingo (Julia Emily Software, France)
- LingTools (Sietec Systemtechnik, Germany)
- LogiTerm (Terminotix Inc., Canada)
- MoBiDic (MorphoLogic, Hungary)
- MTX™ (LinguaTech, USA)
- MultiTerm (TRADOS GmbH, Germany)
- SDL TermBase (SDL, UK)
- TermStar (STAR AG, Switzerland)
- Termwatch (ATRIL Software SL, Spain)
- UniTerm (Acolada GmbH, Germany)
- Xerox Terminology Suite (XTS) (Xerox Multilingual Knowledge Management Solutions, France)

2.3.2 Term extraction tools

Term extraction tools are used to help in setting up terminology. Term extraction tools typically provide a list of potential terms, “term candidates”, from a corpus or from a text, to be validated by a human user.

A term extraction tool may be used when a term base for a new domain has to be developed, or as one of the preparatory steps in a translation workflow.

Term extraction can be either monolingual or bilingual. Monolingual term extraction provides a list of term candidates from a selected corpus or from a text as mentioned above. Bilingual term extraction works on parallel texts, i.e. source texts with their target translation, and identifies the term candidates in the source text and their equivalents in the target text. This means that the first step, identification of the terms in the source text, is the same as for monolingual term extraction.

There are two main approaches to automatic term extraction, linguistic and statistical:

- Linguistic approaches make use of morphologic, syntactic or semantic information. They typically attempt to identify word combinations that match certain part-of-speech patterns, e.g. “adjective + noun”, “noun + ‘de’ + noun”, “noun + noun”, “noun + noun + noun”. So what is involved is simple, or shallow, analysis of the text. Obviously, in order to be able to recognize the nouns etc, a dictionary is needed, and the rules for word formation for terms will be language specific too. A list of stop words, i.e. words that cannot appear as part of a term candidate may be used to avoid some of the mistakes.
- Statistical approaches basically attempt at identifying lexical items or combinations of lexical items that occur with a frequency higher than normal in the corpus. A statistical approach can obviously only be used if a reasonably large corpus is available, and will not work very well for a single text. The advantage of a statistical approach is that it is language independent; however, it also has disadvantages: a purely statistical approach is normally not very satisfactory, i.e. it does not find all the terms and/or it suggests too many candidates that are not terms.

Often the best solution is found by combining the two approaches in a hybrid solution, e.g. a statistical approach followed by a linguistic filtering.

The following table contains examples of term extraction tools:

Names	Descriptions	Addresses
TRADOS Term Extract	Both monolingual and bilingual term extraction. The bilingual part requires access to a bilingual TM. Supports all European languages (Unicode).	http://www.translationzone.com/product.asp?ID=100
Comprendium Terminologist	Provides both monolingual and bilingual term extraction. The bilingual part requires access to a bilingual TM. Currently supporting English, German, French, Spanish.	http://www.comprendium.com/jahia/Jahia/site/lingua/lang/en/pid/448
SDLPhraseFinder	Provides bilingual term extraction. Requires access to parallel texts.	http://www.sdl.com/products-translation/products/sdlphrasefinder-desktop.htm
Xplanation	Offers a service of term extraction, but they seem not to sell the tool.	www.xplanation.com

2.4 Terminology standardization

This chapter provides information on ISO standards that are applicable in the field of terminology, as well as looks at the standardization work accomplished in Germany as a case study.

2.4.1 ISO standards in the terminology field

ISO is a global network of the national standards institutes of 151 countries, on the basis of one member per country, with a Central Secretariat in Geneva, Switzerland, that coordinates the system. The aim of the network is to identify what international standards are required by business, government and society, develop them in partnership with the sectors that will put them to use, adopt them in transparent procedures based on national input, and deliver them to be implemented worldwide.

ISO's work programme ranges from standards for traditional activities, such as agriculture and construction, through mechanical engineering, manufacturing and distribution, to transport, medical devices, the latest in information and communication technology developments, and to standards for services. New growth areas in the coming years are the environment, the service sectors, security and good managerial and organizational practice.

The ISO members propose new standards, participate in their development and, in collaboration with the secretariat, provide support for the 3000 technical committees and subcommittees that actually develop the standards.

The technical committee responsible for the development of standards for terminology is TC 37 – Terminology and other language and content resources. The scope of this technical committee is the standardization of principles, methods and applications relating to terminology and other language resources and content resources in the contexts of multilingual communication and cultural diversity. It consists of four subcommittees which are listed below together with the standards developed by each subcommittee:

ISO subcommittees	Standards
TC 37/SC 1 - Principles and methods	ISO 704:2000: Terminology work – Principles and methods ISO 860:1996: Terminology work – Harmonization of concepts and terms ISO 1087-1:2000: Terminology work – Vocabulary – Part 1: Theory and application
TC 37/SC 2 – Terminographical and lexicographical working methods	ISO 639-1:2002: Codes for the representation of names of languages – Part 1: Alpha-2 code ISO 639-2:1998: Codes for the representation of names of languages – Part 2: Alpha-3 code ISO 1951:1997: Lexicographical symbols and typographical conventions for use in terminography ISO 10241:1992: International terminology standards – Preparation and layout ISO 12199:2000: Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet ISO 12615:2004: Bibliographic references and source identifiers for terminology work ISO 12616:2002: Translation-oriented terminography ISO 15188:2001: Project management guidelines for terminology standardization

TC 37/SC 3 – Systems to manage terminology, knowledge and content	ISO 1087-2:2000: Terminology work – Vocabulary – Part 2: Computer applications ISO 12200:1999: Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange ISO 12620:1999: Computer applications in terminology – Data categories ISO 16642:2003: Computer applications in terminology – Terminological markup framework
TC 37/SC 4 - Language resource management	This subcommittee has not yet developed any standards but several standards are under development

The following sections provide description of some of the most relevant standards concerning terminology.

2.4.2 ISO 704: Terminology work – Principles and methods

This standard gives definitions of all the pivot concepts within terminology such as objects, concepts, concept relations, concept systems, definitions and designations, i.e. it constitutes the foundation of all basic terminology work.

Firstly, objects and concepts are defined. Concepts are divided into individual and general. Then the characteristics (essential, delimiting) of a concept are explained, defining intension and extension. Concept relations are described thoroughly, divided into hierarchical, of which there are two kinds: generic and partitive, and associative, and the nature of concept systems including advice on how to develop them is also treated.

Subsequently, definitions are handled. They are divided into intensional and extensional and there is an exposition of the principles for writing definitions. Key words are reflecting the concept system, conciseness and avoiding deficient definitions (circular, incomplete, negative). Notes for secondary information and graphic representations are introduced.

Designations are the representation of a concept in natural language. Designations can be terms (one or more words) that designate general concepts, appellations that designate individual concepts and symbols that designate both individual and general concepts. The standard explains homonymy, synonymy and treats the general principles for term formation.

Finally, there is a brief introduction to some standardization issues, explaining preferred, admitted and deprecated terms and the reasons for deprecation of terms.

An appendix gives examples of term-formation methods in English, i.e. new forms formed e.g. by derivation, compounding or abbreviation, and existing forms transformed into terms e.g. by changing the syntactic category.

2.4.3 ISO 860:1996 Terminology work – Harmonization of concepts and terms

Harmonization of terminology is a highly relevant issue for both monolingual and bi- or multilingual communication since the aim is to minimize the terminological difficulties of communication. This standard handles the process of harmonization of terms and concepts from the preliminary analyses of subject fields and concept systems to the construction of harmonized concept systems, definitions of harmonized concepts and harmonization of the individual terms.

Before a concept harmonization can be carried out, several analyses have to take place. An analysis of the subject field is obligatory since subject fields that are well established, with a tradition of standardization or which deal with concrete objects are more likely to result in successful harmonization than subject fields under development, or subject fields within humanities or social sciences and with no tradition for standardization. If the chances for a successful harmonization seem good, a preliminary analysis of the concepts and their characteristics has to be carried out.

The harmonization process is as follows:

- It starts with a comparison of the involved concept systems in terms of the number of concepts, relations between concepts, depth of structure and type of characteristics leading to the construction of harmonized concept systems.
- All the concepts must then be analysed by comparing the definitions. If the definitions differ, it must be decided whether the difference is relevant or irrelevant. If relevant, it means that there are indeed two or more different concepts involved that must be defined and placed in the concept system.
- The defining characteristics for the harmonized concepts have to be established.
- When the concepts are harmonized, the terms can be harmonized taking into account the differences and similarities between languages, the tradition of term formation in the subject field and in a given language as well as the already established terminology.

The standard includes a flow chart of the preliminary analyses and the harmonization process.

2.4.4 ISO 10241: International terminology standards – Preparation and layout

This standard gives a practical introduction on how to write an international standard for terminology. The aim of an international standard on terminology is the harmonization of concepts, concept systems and terms in a given subject field in different languages. This standard deals with the procedure for developing standards whereupon some principles for terminography are given.

The preliminary work consists of defining the target group, delimiting the subject field defining the scope and the sub-fields and analyzing the terminological usage of the selected sources (incl. evaluation of sources). The work should be carried out simultaneously for all languages involved and the number of terms dealt with should be limited.

All possible terminological data should be collected and recorded, extracting all relevant material for term lists, concepts and definitions in one operation from the source data. From this material, first the term lists for each language should be established after which the concepts and their relations should be specified establishing concept systems. Finally a comparison and harmonization of the language specific concept systems should be performed. The last part of the working procedure is to write the definitions in accordance with ISO 704.

The terminography part goes through the important data categories. The essential information is entry number, preferred term and definition; additional information comprises among other things symbols, pronunciation, grammatical information, subject field, references, examples of usage and term equivalents. The preferred order of the entries is systematic order, but alphabetical order as well as mixed order is possible. Alphabetical standards shall contain systematic indexes and vice versa.

2.4.5 ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange

ISO 12200 is an international standard for the interchange of terminological data allowing the distinct identification of separate data sets and data categories as well as its dependencies and relations. MARTIF has been elaborated in order to facilitate more universal, less costly exchange of data collections containing concept-oriented terminology entries. The format relies heavily on the data category names and definitions contained in the companion standard ISO 12620. MARTIF is based on Standard Generalized Markup Language (ISO 8879, SGML) and was originally developed in close cooperation with the Text Encoding Initiative (TEI) and the Localization Industry Standards Association (LISA).

As an SGML-based solution, MARTIF has the additional advantage that terminological data can be easily processed like any other SGML document, e.g. for the publication of printed terminological glossaries. Due to its high degree of flexibility MARTIF is able to adequately represent all forms and structures of terminology resources.

MARTIF not only provides an open, flexible mechanism for exchanging data with other potential users employing different terminology management systems. It can also be used when companies need to change or upgrade software from one database format to another.

The main body of the MARTIF standard specifies the formalism to be used in preparing terminology data collections for interchange by defining the SGML Document Type Definition (DTD) and listing the appropriate tags (markup) used to structure the data. Normative Annex A of the standard specifies the markup for the individual terminological data categories to be used in the MARTIF environment, based on ISO 12620.

A complete MARTIF document consists of a prolog, followed by a document instance of type MARTIF. The document instance consists of a <martifHeader> followed by the text, which in turn consists of optional front matter, the <body> (a sequence of terminological entries), and optional <front> and <back> matters.

The following code sample shows the basic components of a MARTIF document:

```
I. Prolog
II. Document instance (<martif lang=en>)
     A. header (<martifHeader>)
     B. text
          1. front (optional)
          2. body
               a. first terminological entry <termEntry> (minimum of one)
               b. second terminological entry <termEntry>
               c. etc. (additional terminological entries)
          3. back (optional)
```

The following code sample shows the structure of the document instance:

```
<martif lang=en>
<martifHeader>
... (The header goes here.)
</martifHeader>
<text>
<body>
... (The terminological entries go here.)
</body>
<back>
... (Included bibliographic entries go here.)
... (Any external references (<xref>s) also go here.)
</back>
</text>
</martif>
```

The following code sample shows an example of a full MARTIF term entry:

```
<termEntry id=’ID000073578’>
    <descrip type=’subjectFieldLevel1’> appearance of materials </descript>

    <ntig lang=en>
        <termGrp><term>opacity</term>
        <termNote type=’partOfSpeech’>n</termNote></termGrp>
        <descripGrp><descrip type=’definition’>degree of obstruction to the transmission of
        visible light</descrip><ptr type=’sourceIdentifier’ target=’ASTM.E284’></descripGrp>
        <adminGrp><admin type=’responsibility’>E12</admin> </adminGrp>
    </ntig>

    <ntig lang=de>
        <termGrp><term> Opazit&auml;t</term>
        <termNote type=’partOfSpeech’>n</termNote>
        <termNote type=’gender’>f</termNote></termGrp>
        <descripGrp><descrip type=’definition’>Ma&szlig; f&uuml;r die
        Lichtundurchl&aumlt;ssigkeit </descrip><ref type=’sourceIdentifier’ target=’DIN-
        6730.1996-05’>p. 383</ref></descripGrp>
        <adminGrp><admin type=’responsibility’>Normenausschu&szlig; Papier und Pappe
        (NPa) im DIN Deutsches Institut f&uuml;r Normung e.V. </admin></adminGrp>
    </ntig>
    <ntig lang=fr>
        <termGrp><term>opacit&eacute;</term>
        <termNote type=’partOfSpeech’>n</termNote>
        <termNote type=’gender’>f </termNote> </termGrp>
        <descripGrp><descrip type=’definition’>rapport du flux lumineux incident au flux
        lumineux transmis ou r&eacute;fl&eacute;chi par un noircissement
        photographique</descrip>
        <ptr type=’sourceIdentifier’ target=’HJdi1986’></descripGrp>
        <adminGrp><admin type=’responsibility’>C.I.R.A.D.</admin> </adminGrp>
    </ntig>
</termEntry>
```

As noted above, MARTIF was originally designed for the so-called negotiated interchange, where partners examine each other's data before interchange and make decisions about pre-conditioning the data before importing it from the interchange format. This approach allows a high degree of flexibility in individual applications.

For a more "blind" interchange, specific MARTIF-compatible formats can be defined on the basis of ISO 16642. Following the structure of MARTIF the XML-based standard format for terminological data TermBase eXchange (TBX) has been developed.

2.4.6 ISO 12620:1999 – Computer applications in terminology – Data categories

Terminological data are collected, managed, and stored in a wide variety of environments. For purposes of storage and retrieval, these data are organized into terminological entries, each of which traditionally treats information associated with a single concept. Data items appearing in individual terminological entries are themselves identified according to data category. Differences in approach and individual system objectives inevitably lead to variations in data category definition and in the assignment of data category names. The use of uniform data category names and definitions, at least at the interchange level, contributes to system coherence and enhances the reusability of data.

The International Standard ISO 12620:1999 specifies data categories for recording terminological information in both computerized and non-computerized environments and for the interchange and retrieval of terminological information independent of the local software applications or hardware environments in which these data categories are used.

The data category specifications are divided into three major groups: data categories for terms and term-related information, descriptive data, and administrative data. The groups are further subdivided into ten sub-groups.

Term and term-related data categories:

- Subgroup 1 consists of the data category term and contains a term or other information treated as if it were a term (e.g., phraseological units and standard text).
- Subgroup 2 specifies data categories for term-related information.
- Subgroup 3 specifies data categories for information relating to equivalence between or among terms assigned to the same or very similar concepts.

Descriptive data categories:

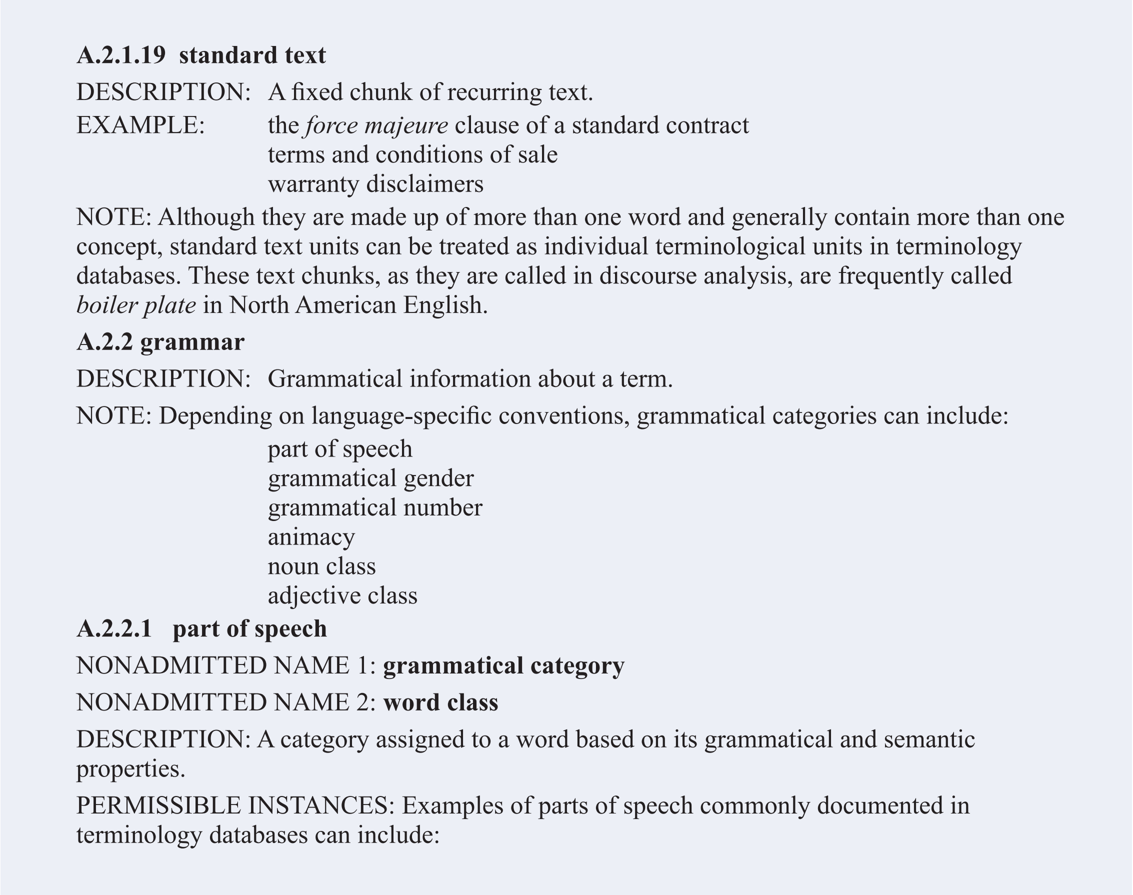
- Subgroup 4 specifies data categories for the classification of concepts into subject fields and subfields, along with other classification-related information.
- Subgroup 5 specifies data categories for concept-related description, i.e., different kinds of definitions, explanations and contextual material provided to define or otherwise determine the subject field and concept to which a term is assigned.
- Subgroup 6 specifies data categories for indicating relations between pairs of concepts.
- Subgroup 7 specifies data categories used to express the position of concepts within concept systems.

- Subgroup 8 specifies the data category note. This category stands alone because it can be associated with any one of the other categories and therefore cannot be subordinated to any other specific subgroup.

Administrative data categories:

- Subgroup 9 specifies data categories for documentary languages and thesauri.
- Subgroup 10 specifies data categories for all other strictly administrative information.

The following illustration is an example for the description of data categories in ISO 12620:



A.2.1.19 standard text

DESCRIPTION: A fixed chunk of recurring text.

EXAMPLE: the *force majeure* clause of a standard contract
terms and conditions of sale
warranty disclaimers

NOTE: Although they are made up of more than one word and generally contain more than one concept, standard text units can be treated as individual terminological units in terminology databases. These text chunks, as they are called in discourse analysis, are frequently called *boiler plate* in North American English.

A.2.2 grammar

DESCRIPTION: Grammatical information about a term.

NOTE: Depending on language-specific conventions, grammatical categories can include:
part of speech
grammatical gender
grammatical number
animacy
noun class
adjective class

A.2.2.1 part of speech

NONADMITTED NAME 1: **grammatical category**

NONADMITTED NAME 2: **word class**

DESCRIPTION: A category assigned to a word based on its grammatical and semantic properties.

PERMISSIBLE INSTANCES: Examples of parts of speech commonly documented in terminology databases can include:

Figure 6: Example of data categories description in ISO 12620.

If applied for the purpose of interchanging machine-readable terminology, it is recommended that this standard is used in conjunction with ISO 12200, although it can also be used for modelling terminological information independently of computer applications.

The data categories specified in ISO 12620 constitute the basis for various other standards dealing with the processing of terminological data, e.g. ISO 12200. The data categories correspond to data element concepts in the ISO/IEC 11179 series of standards.

International Standard ISO 12620 was prepared by Technical Committee ISO/TC 37, Terminology (principles and coordination), now called Terminology and other language resources, Subcommittee SC 3, Computer applications for terminology.

At present TC 37 is revising ISO 12620 and planning to publish the new version entitled Computer applications in terminology — Data categories — Model for description and procedures for maintenance of data category registries for language resources. This revision implicates quite fundamental changes in the handling of data categories, meaning that in the future, the data categories will be maintained in a universal Meta Data Registry available in the Internet, whereas ISO 12620 will provide all the specifications indispensable for the description of data categories, as well as the admission procedure for new data categories to be added to the Data Category Registry. In this context, the foundation of a new subcommittee of ISO/TC 37, SC 4 Language resource management, will grant access to collections of data categories beyond those defined by SC 3, e.g. the language codes defined in ISO 639-1 and ISO 639-2.

2.4.7 ISO 16642:2003 – Computer applications in terminology – Terminological markup framework (TMF)

This international standard has been developed to facilitate the use and re-use of terminological data collections, taking into account the real-life conditions of different formats, database environments and term-bank systems as well as the various data models the collections are based on. The standard is also motivated by the need to provide better connections between terminological databases and other lexical resources dedicated, for instance, to machine translation or natural language processing.

The core element is a single high-level meta model representing a unique information structure shared by all terminology mark-up languages (TML), which decomposes the organization of a terminological database into basic components – the structural skeleton, defined as a "set of XML elements which, in a given TML, results from the expansion of the meta-model", and the elementary units of information (i. e. data categories) that can be attached to the structural skeleton.

For mapping any given format, or TML, onto the abstract components of TMF, a simplified XML application has been defined. This format is called GMT (Generic Mapping Tool) and is based on a reduced set of XML elements and attributes, which serve as containers for nodes of the structural skeleton (identified by <struct> tags) and data categories (identified by <feat> tags).

Thus, the data of a terminology data base expressed in any format is mapped onto a given data model using GMT, by

- decomposing every entry into the three structural levels of the meta-model, the Terminological Entry (TE), the Language Section (LS) and the Term Section (TS) (<struct> element); and
- expressing each information unit by means of the <feat> element where the type signifies the data category name.

To illustrate how a terminological data collection can be analysed as an abstract structure, a simple terminological entry expressed as an XML document conformant to MSC (MARTIF with Specified Constraints, a variant of ISO 12200) specifications is mapped via the TMF meta-model onto data categories defined in ISO 12620.

Terminological entry as expressed in MSC:

```
<?xml version=”1.0”?>
<martif type=”MSC” lang=”en”>
 <text>
 <body>
  <termEntry id=”ID67”>
  <descrip type=”subjectField”>manufacturing</descrip>
  <descrip type=”definition”>
   A value between 0 and 1 used in ...
  </descrip>
  <langSet lang=”en”>
   <tig>
   <term>alpha smoothing factor</term>
   <termNote type=”termType” datatype=”picklistVal”>fullForm</termNote>
   </tig>
  </langSet>
  <langSet lang=”hu”>
   <tig><term>Alfa ...</term></tig>
  </langSet>
  </termEntry>
 </body>
 </text>
</martif>
```

This document can be mapped to the abstract model by identifying a structural skeleton corresponding to the meta-model and by associating the corresponding information units with each structural node in the structural skeleton, as shown below:

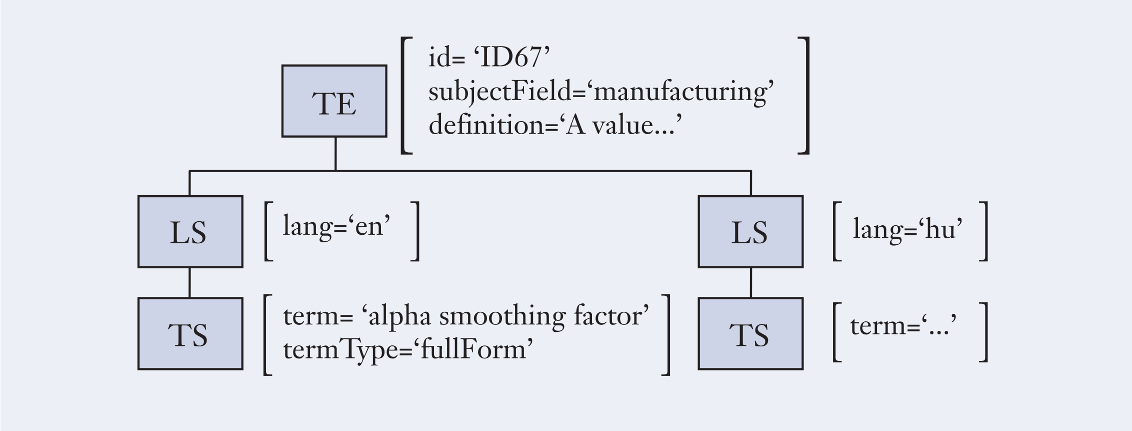


Figure 7: Mapping of the term entry to the abstract model.

The data categories correspond to the data categories specified in ISO 12620 as follows:

Data category	ISO 12620 number	ISO 12620 name
Id	A10.15	entry identifier
subjectField	A04	subject field
definition	A05.01	Definition
Lang	A10.07.01	language identifier
Term	A01	Term
termType=fullForm	A02.01.07	full form

One possible encoding in the GMT format is shown below:

```
<?xml version=”1.0” encoding=”iso-8859-1”?>
<tmf>
  <struct type=”TE”>
     <feat type=”entry identifier”>ID67</feat>
     <feat type=”subject field”>manufacturing</feat>
     <feat type=”definition”>A value between 0 and 1 used in ...</feat>
     <struct type=”LS”>
       <feat type=”language identifier”>en</feat>
       <struct type=”TS”>
         <feat type=”term”>alpha smoothing factor</feat>
         <feat type=”term type”>fullForm</feat>
       </struct>
     </struct>
     <struct type=”LS”>
       <feat type=”language identifier”>hu</feat>
       <struct type=”TS”>
         <feat type=”term”>Alfa ...</feat>
       </struct>
     </struct>
   </struct>
</tmf>
```

The combination of the meta-model and a given Data category specification (DCS) is enough to define the degree of interoperability of two TMLs, encompassing its full informational properties from a terminological point of view. Any information structure that corresponds to such conditions has a canonical expression as an XML document using the GMT (Generic Mapping Tool) representation. The interoperability between two different TMLs depends solely on their compatibility at that level. The illustration below shows the interoperability between two TMLs using GMT:

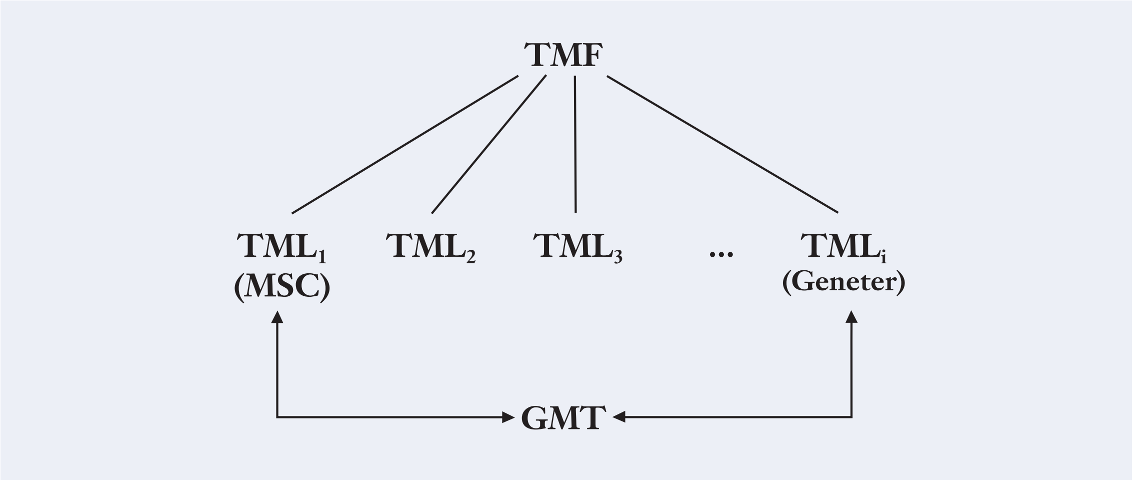


Figure 8: Interoperability between two TMLs using GMT.

When two TMLs are based upon two different DCSs, GMT provides a framework for identifying what information can be transformed between one format and another and what will be lost during the transformation.

The comparison between two TMLs is only possible if there is a central repository of data categories, associated with a consistent model for these, which can act as a broker between any two formats. For the application of this standard, ISO 12620 forms a reference Data Category Registry (DCR) for any information unit to be used in the definition of a TM.

2.4.8 Terminology standardization in Germany: a case study

DIN, the German Institute for Standardization (Deutsches Institut für Normung e. V.) was founded in 1917. Its head office is in Berlin. Since 1975 it has been recognized by the German government as the national standards body and represents German interests at international and European level.

DIN offers a forum in which representatives from the manufacturing industries, consumer organizations, commerce, the trades, service industries, science, technical inspectorates, government, in short anyone with an interest in standardization, may meet in order to discuss and define their specific standardization requirements and record the results as German Standards.

Standardization as undertaken by DIN is a service that aims to benefit the entire community. The results of its work have a significant influence on economic performance at both company and national level. A research project completed in 2000 confirmed the annual benefit to the German economy being 1 % of GNP, or approx. US$ 15 billion.

DIN Standards promote rationalization, quality assurance, safety, and environmental protection as well as improve communication between industry, technology, science, government and the public domain.

The main activity of DIN is the development of technical rules. In drawing up a new organizational structure, a clear line has been drawn between standardization on the one side and business activities on the other. The objective of DIN is to create standards for the benefit of the economy and the society as a whole. The business activities of the companies within the DIN Group are profit-oriented. The income generated by the subsidiary companies of DIN, and from those companies in which it is a shareholder, represents the single largest contribution to the financing of the not-for-profit core activities of DIN.

The input of external experts into standardization is organized in standards committees and their subsidiary working bodies. One standards committee is responsible for each distinct area of activity and also coordinates the corresponding standardization work at European and international level. As a rule, the standards committee in DIN comprise a number of technical committees. There are currently 76 standards committees, in which some 26,000 external experts are working as voluntary delegates on the standards. Draft standards are published for public comment, and all comments are reviewed before final publication of the standard. Published standards are reviewed for continuing relevance every five years, at least. In 2004, the number of DIN standards amounts to nearly 29,000, from which 15,200 are available in English. In 2004, DIN achieved a turnover of 56 million EUR.

The standards are published and sold by the publishing house Beuth Verlag that specializes in sales and distribution of standards, directives and other normative documents.

Within terminology standardization two main directions can be distinguished: the standardization of concepts and the respective terms and the definition of principles and guidelines for terminology work and terminology standardization. The standardization of concepts and terms is in general carried out by subcommittees of the respective technical Standards Committees. They include the standardized terminology in the corresponding technical standard or publish it in the form of specific terminology standards.

The concepts enclosed in DIN standards with their terminological representations, definitions and further information are documented in a terminology database called DIN-TERM. As the standards often contain foreign-language terms and definitions, DIN-TERM covers not only German terms, but also English and French equivalents. The service responsible for DIN-TERM plans to make the data available to users in the form of technical dictionaries and electronic databases. Today DIN-TERM contains more than 210,000 entries.

For the standardization in the field of the principles of terminology work the Terminology Standards Committee (Normenausschuss Terminologie, NAT) is responsible. It focuses on the fundamental significance of technical language for standardization in general as well as on tools for terminology work, translation and lexicography. The primary areas of responsibility are: principles of concept and term formation, elaboration and configuration of technical dictionaries, computer applications for terminology work and lexicography, terminology of the terminology work, terminological practice and technical translation. NAT represents German interests in ISO/TC 37 Terminology and other language resources.

CHAPTER 3

LEGAL FRAMEWORK FOR TERMINOLOGY DATABASES

This chapter deals with the ever important and sensitive legal aspects of terminology resources. It provides an overview on copyright issues and applicable EU legislation, as well as includes the Infoterm Code of Good Practice in Copyright.

3.1 Overview on copyright questions

The Berne Convention for the Protection of Literary and Artistic Works, adopted at Berne in 1886 and revised since then several times, constitutes the basis for copyright law in all countries which notified it. In the course of implementation into national legislations, however, the stipulations of the Berne Convention undergo subtle differences. Nevertheless it provides minimum standards of protection, such as

- the right to translate, the right to make adaptations and arrangements of the work,
- the right to perform in public dramatic, dramatico-musical and musical works,
- the right to recite in public literary works,
- the right to communicate to the public the performance of such works,
- the right to broadcast (with the possibility of a contracting State to provide for a mere right to equitable remuneration instead of a right of authorization),
- the right to make reproductions in any manner or form (with the possibility of a contracting State to permit, in certain special cases, reproduction without authorization provided that the reproduction does not conflict with the normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author, and with the possibility of a contracting State to provide, in the case of sound recordings of musical works, for a right to equitable remuneration),
- the right to use the work as a basis for an audiovisual work, and the right to reproduce, distribute, perform in public or communicate to the public that audiovisual work. The application of copyright presupposes an individual intellectual creation of the author.

In Europe, copyright exists from the creation of a work and does not require formal registration or notice (cf. IPR Enforcement Directive, Preamble recital 19). Article 15 of the Berne Convention establishes the presumption whereby the author of a literary or artistic work is regarded as such if his/her name appears on the work. Whereas the moral rights of an author are not transferable, the exploitation rights of the work may be granted to a third party by the author.

In nearly all countries copyright is subject to limitations and exemptions for the public benefit allowing the reuse of data without special permission or payment of royalty fees, for example for private purposes, for the purposes of illustration for teaching or scientific research – as long as the source is indicated – or for the purposes of public security or an administrative or judicial procedure. Whereas in most European legal systems these limitations and exemptions are enumerated explicitly, the Anglo-American law contains the more general "fair use", or "fair dealing" doctrine (cf. Wright 1996, http://en.wikipedia.org/wiki/Fair_use 27.10.2005).

Modern information and communication technologies – allowing hitherto unimaginable ways and means of replication and conversion – have put copyright provisions under stress. Furthermore, under a mobile content (mContent) perspective, today, copyright is increasingly impacted. mContent – including terminology – has to be considered from the outset as:

- multilingual
- multimodal
- multimedia

They should be prepared in such a way that it meets multi-channel and universal accessibility requirements, comprising also the requirements of people with special needs. This is the beginning of a new proliferation of derivative works, which may need further modifications or extensions of the Berne Convention.

3.1.1 Copyright in terminology

When it comes to terminology, experts hold quite different views on the question in which way, if at all, terminology as such and terminology collections are subject to copyright or other intellectual property rights. The opinion that the creation of new technical terminology and the formulation of definitions and concept descriptions should be considered as a creative mental achievement worth protection by copyright is rather uncontested by subject-field and terminology experts in general. However, it is contested by legal experts based on the fact that terminological data represent the state-of-the-art, which does not qualify them as original work (cf. RaDT 2000). The question whether the compilation of terminology is also subject to copyright must be decided apparently depending on the character of the compilation (cf. Budin 1993).

The Guide to Terminology Agreements by Infoterm states, that “while concepts, as ‘units of knowledge’, should be regarded as the intellectual property of all mankind, their representations as terms and definitions, or other kinds of concept description, as graphic symbols, or as other kinds of non-linguistic representation must be considered to be the intellectual property of the originator, i.e. a single expert, group of experts, or institution/organization, if this information has been conceived or prepared by the respective originator in the form of a terminological entry, a specific sub-section of an entry, or a collection of terminological data” (Guide 1996, cf. Annex 4) – whereas other experts deny that a single term or terminological entry is already subject of copyright in any case (cf. Stellbrink 1993, p. 4). It seems to be obvious that the character of every entry determines the question whether it can be considered as a creative work and subject to copyright (cf. Goebel 1993, p. 41).

However, while the “legal position with regard to the definition of the ‘smallest unit’ that may be asserted on the bases of the protection of intellectual property, is not yet settled” (GTW-Report 1996, p. 28), a terminology database is covered by copyright in Europe as a sui generis right granted by the EU Database Directive, because the compilation and presentation of the data has to be considered as an autonomous creative work independently from its content and, indeed, often in addition to its copyrighted components. Protected by this right is a database as a whole or a “substantial part” (Database Directive, Art. 7(1)) of it.

A complex terminology database in general consists of linguistic and non-linguistic knowledge representations, and may contain names and logos being part of a term or concept description. The first type comprehends primarily:

- terms proper, incl. abbreviations, nomenclature names, etc.;
- terminological phraseology;
- thesaurus descriptors and class names of a subject classification scheme;
- definitions and other kinds of textual concept description;
- statements representing a (micro-)proposition, and contexts or co-texts.

Non-linguistic knowledge representation can appear in form of:

- formulae, e.g. in mathematics and chemistry;
- alphanumeric codes, or equivalents, such as barcodes;
- graphical symbols, complex graphs, figures, and images.

In a new multimedia encyclopaedia, moving animations or pictures, sound, video clips and other kind of representation can be found, which sooner or later will also find their way into terminology databases.

Terminology databases can consist of different database files – terminology files, bibliographic data files, and indexing and retrieval language files. Terminology files contain the entries whose data consolidate around the term and concept description as the most important elements to represent the concept in question. Bibliographic data files consist of entries containing references, such as author, title, year of publication, abstract, etc. and other sources of terminological information. In the terminology entries themselves source-related information will occur in coded form, rather than as a full bibliographic entry. However, each code points to a full bibliographic record mostly stored in a separate bibliographic file. Indexing and retrieval language files comprehend records containing class names or thesaurus descriptors, subject headings, etc. for indexing – the terminological records as well as bibliographic records – and retrieval purposes.

In order to analyze the copyright situation with regard to terminology databases the national legislation of the respective country is authoritative. The regulation in all EU member states is similar, because it is based on the relevant EU Directives implemented into the national legislation of each of them. This does not pose any difficulty as long as a copyrighted work is solely used, first of all in printed form, in a given country. The problems arise, if a copyrighted work is accessible via the Internet – i.e. globally.

The identification of the copyright holder may pose several problems: different types of data, such as textual information and pictures, can be subject to different copyrights which can be owned by different people, i.e. authors/originators being natural persons. The copyright for the content of a terminology database can be owned by one or several persons, when the data collection is the result of a collective work. Big terminology databases may comprise several collections subject to the copyright of different groups of people, and the database as such can constitute an additional copyright for those who created the database.

Although the moral copyright of the author of terminological records in a terminology database remains with the author in any case, the author can authorize another natural or legal person to use the data, in particular to reproduce, modify, translate, and distribute it according to the exploitation rights. The sui generis right based on the Database Directive can belong to the person(s) or legal person(s) that have established the database.

For the purposes of EuroTermBank, after selecting appropriate terminology resources and identifying their copyright holders, the Consortium signs agreements with them in order to obtain the right to use the data for the EuroTermBank portal. These agreements entitle the Consortium to modify the data according to formal and technical requirements, and to digitize data available only as hard copy.

The conditions or restrictions of use defined by copyright holders have to be taken into account. They can refer to the price, in the case of data available only for a fee, and payment procedures, or to any restrictions of the use and distribution of the data.

Eventually the measures appropriate to prevent misuse and to guarantee due acknowledgment of the authors have to be investigated. These preconditions determine the design of the agreements to be concluded with the copyright holders.

3.1.2 Groups of copyright holders

For the approach to copyright holders and the design of contracts to be signed, the possible interests of the copyright holders should be taken into account. There are two main groups of copyright holders (holders of exploitation rights) pursuing, in general, various targets: authors and publishers.

Authors of terminology collections may be single persons or groups of authors, institutions like professional organizations or standardization boards, or, at least, companies creating their own terminology. If they are legal persons, the actual authors/originators as a rule have ceded to them – by blanket agreement or any other kind of implicit or explicit agreement – the exploitation rights. In fact, an author can strive for very different benefits:

- He/she could be interested in the first place in remuneration or some other economic effect.
- He/she may as well wish to disseminate his/her work widely in order to make it available and useful to a group of users as large as possible.
- He/she could, in his/her capacity as an expert of his/her subject field, have an ethical interest in fostering the standardization of terminology in this field, in contributing to avoid duplication of work in order to achieve a harmonized mono- and multilingual technical communication and, finally, to enhance scientific, technical and economical co-operation.

Hopefully, authors appreciate the scientific dialogue and desire to cooperate with EuroTermBank, and, thus, are ready to get involved in an important project using state-of-the-art technology and methodology.

The second large group of copyright holders (holders of exploitation rights) consists of the publishers. They are vendors of terminology, so their natural interest is to sell the data to the user via EuroTermBank. There are cases when the authors are also the publishers of their terminology collection.

A subset of the two first groups of copyright holders are the standards developing organizations (SDOs), which due to their importance for the creation of reliable terminology and to their special function are described here as a single group. As a rule terminologies in SDOs are created by groups of experts in the form of technical committees, their sub-committees or working groups. By the very nature of the SDOs, their terminologies must be considered as authoritative data.

However, whether implicitly or stated explicitly, the terminological data are strictly speaking only valid within the scope of the respective committee or of the standard developed by them. Nevertheless these data are of crucial importance to the process of development of subject standards, especially for their quality. Because of traditional working methods, standardized terminologies are more often than not recorded in conventional form or in an electronic format not appropriate for data processing. Given their authoritative nature SDOs consider standardized terminologies as one of their special assets, although in many cases they lack the means and skills to market them. One of the reasons for this may be that the general business model for standardization does not fit as a business model for standardized terminologies. Here EuroTermBank could step in and provide a valuable platform for the distribution of standardized terminologies on a commercial or non-commercial basis. While a general agreement in this regard has not been obtained for the time being, a long-term cooperation with international and national SDOs is considered as highly desirable.

3.2 EU legislation on copyright and related rights

The following EU Directives are of relevance to the ETB project:

- Directive 91/250/EC of 14 May 1991 on the legal protection of computer programs;
- Directive 92/100/EC of 19 November 1992 on rental right and lending right and on certain rights related to copyright in the field of intellectual property;
- Directive 93/83/EC of 27 September 1993 on the coordination of certain rules concerning copyright and rights related to copyright applicable to satellite broadcasting and cable retransmission
- Directive 93/98/EC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights
- Directive 96/9/EC of 11 March 1996, on the legal protection of databases;
- Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market
- Directive 2001/29/EC of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society.
- Directive 2004/48 EC of 29 April 2004 on the enforcement of intellectual property rights.

For the purpose of the EuroTermBank project, mainly four EU Directives provide for the legal framework of intellectual property rights and copyright issues:

- Directive 96/9/EC of 11 March 1996 on the legal protection of databases
- Directive 2001/29/EC of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society
- Directive 2000/31/EC of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market
- Directive 2004/48 EC of 29 April 2004 on the enforcement of intellectual property rights

They are already implemented into national law by the EU member states. Although they grant options to the member states' legislators, they aspire to harmonize the EU-wide legislation on IPR and copyright, taking into account the new technologies offering a wide range of possibilities for creating, storing, reproducing and distributing intellectual works.

Directive 96/9/EC, called the Database Directive, stipulates the harmonization of copyright provisions in the member states, and provides for a new sui generis right entitling the author of a database under certain conditions to prevent extraction and/or re-utilization of the whole or of a substantial part of the contents of his database.

In this connection it must be repeated that the copyright for the content of a terminology collection can be owned by one or several authors, when the data collection is the result of a collective work. Big terminology databases may comprise several collections subject to the copyright of different groups of authors, and the database as such can constitute an additional copyright for those, who created the database. The moral copyright of the author – being a natural person – of a terminology database remains with him in any case, but the author can authorize another natural or legal person to use the data, in particular to reproduce, modify, translate, and distribute it according to exploitation rights.

Directive 2001/29/EC, called the European Union Copyright Directive (EUCD), covers mainly three areas, considered as crucial for information in cyberspace. It grants authors with a new exclusive right to communicate their works to the public, it deals with copyright limits, i.e. exceptions and limitations to the reproduction right for digital works, and it provides legal protection for technical measures dedicated to safeguard rights.

Directive 2000/31/EC, called the E-commerce Directive, intends to "improve the legal security of such commerce in order to increase the confidence of Internet users. It sets up a stable legal framework by making information society services subject to the principles of the internal market (free circulation and freedom of establishment) and by introducing a limited number of harmonized measures" (http://europa.eu.int/scadplus/leg/en/lvb/l24204.htm, 03.12.2005).

Directive 2004/48/EC aims at providing an "equivalent level of intellectual property protection throughout the whole European Community" (IPR Helpdesk 2005 p. 1). It establishes specific legal measures and procedures to be taken in case of infringement of IPR.

3.2.1 Directive 96/9/EC on the legal protection of databases

The goal of the Directive is the legal protection of databases, whereas database is defined as a "collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means" (Database Directive, Art. 1(2)). The definition covers not only electronic, but also paper databases. The Directive is subdivided into two main parts: on the one hand it stipulates the harmonization of copyright provisions in the member states, and on the other hand it provides for a new sui generis right.

The protection by copyright extends to "databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation" (ibid., Art. 3(1)), the content of the databases itself not being subject to regulation by this act. However, the subject matters being protected under copyright or related rights, which are incorporated into a database, remain protected by the respective rights and may not be incorporated into the database without the permission of the copyright holder.

The author of a database, i. e. the natural person who created it, has the exclusive right to carry out or to authorize the reproduction, the translation, adaptation, arrangement or other alteration as well as the distribution, communication, display or performance to the public (cf. ibid., Art. 5). In case of a collective work created by a group of natural persons, the exclusive rights devolve to them jointly.

The Directive entitles the member states to provide exceptions to copyright in case of reproduction for private purposes, for teaching or scientific research, for the purposes of public security, administrative or judicial procedure, or in other cases, traditionally authorized under national law, provided that the copyright holder's legitimate interests are not unreasonably prejudiced, according to the Berne Convention for the protection of Literary and Artistic Work (cf. ibid., Art. 6).

Besides the copyright regulation, the Directive provides a sui generis right for the author of a database "which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database" (ibid., Art. 7(2)), whereas

- "'extraction' shall mean the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form;
- 're-utilization' shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission" (ibid.).

Furthermore, the Directive prohibits the "repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database" (ibid., Art. 7(5)). Public lending is not considered as an act of extraction or re-utilization, though.

Besides the harmonization of rights and obligations of the lawful users of databases, the Directive grants the member states the right to stipulate exceptions to the sui generis right, according to the above mentioned exceptions to copyright regulations.

The sui generis right is limited to fifteen years from the first of January of the year following the date of completion of the making of the database. The Directive (Art. 19(2)) establishes the principle that when a database is substantially changed – to be evaluated qualitatively and/or quantitatively – it becomes a new database, entitled to its own term of protection.

The member states are obligated to provide appropriate remedies for infringements of the rights provided for in the Directive.

3.2.2 Directive 2001/29/EC on the harmonization of certain aspects of copyright and related rights in the information society

The Directive serves to implement international obligations of the Community accepted by signing the WIPO (World Intellectual Property Organization) Copyright Treaty and the WIPO Performances and Phonograms Treaty.

It intends to adapt the existing copyright regulations in order to respond to technological developments, i.e. digital technologies, and economic realities offering new forms of creation, production and exploitation. For a smooth functioning of the European internal market, the various national provisions on copyright and related rights needed to be harmonized.

The Directive extends copyright to digital products and aims at safeguarding the rights and interests of different categories of copyright holders such as authors, performers, phonogram producers, producers of the first fixations of films, or broadcasting organizations.

The Directive stipulates that member states shall provide for authors for the exclusive rights to authorize or prohibit the reproduction of their work in any form, to communicate it to the public and to authorize or prohibit any form of distribution to the public. The distribution right is exhausted when "the first sale or other transfer of ownership in the Community of that object is made by the copyright holder or with his consent" (Copyright Directive 2, Art. 4(2)).

Member states have a significant freedom in the establishment of exceptions and limitations for certain cases, for example educational and scientific purposes, for the benefit of public institutions such as libraries and archives, for purposes of news reporting, for quotations, for use by people with disabilities, for public security uses and for uses in administrative and judicial proceedings. They can also provide for exceptions or limitations to the reproduction right concerning reproduction for private use, accompanied by fair compensation. In most of these exceptional reproduction cases the source, including the author's name, shall be indicated.

The Directive imposes the obligation to provide adequate legal protection against the circumvention of technological measures carried out in order to safeguard the rights of an author or another copyright holder, i.e. to "prevent or restrict acts, in respect of works or other subject-matter, which are not authorized by the copyright holder of any copyright or any right related to copyright as provided for by law or the sui generis right provided for in Chapter III of Directive 96/9/EC" (ibid., Art. 6(3)).

3.2.3 Directive 2000/31/EC on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market

The goal of the Directive is the "proper functioning of the internal market by ensuring the free movement of information society services between the Member States" (E-Commerce Directive, Art. 1(1)). It harmonizes the rules and regulations concerning e-commerce within the internal market in order to provide legal certainty in this area. The principles of freedom to provide services and freedom of establishment shall apply to Information Society services throughout the EU, provided that they comply with the law in the respective member state.

"The Directive covers all Information Society services, both business to business and business to consumer, and services provided free to the recipient (for example funded by advertising or sponsorship revenue). Examples of online sectors and activities covered include shopping, newspapers, databases, financial services, professional services (such as lawyers, doctors, accountants, estate agents), entertainment services, direct marketing and advertising and internet intermediary services" (Press release of the European Commission on http://europa.eu.int, 12.10.2005).

The Directive establishes harmonized rules on mandatory information an online service must provide to users, such as name, address, or contact details, on commercial communications and electronic contracts. It restricts the liability of intermediary service providers concerning the information transmitted, temporary storage of information, and information monitoring. The member states and the Commission shall encourage the establishment of codes of conduct at Community level in order to facilitate the proper implementation of the Directive's provisions.

3.2.4 Directive 2004/48 EC on the enforcement of intellectual property rights

Based on consultations, and despite the TRIPs (Agreement on Trade-Related Aspects of Intellectual Property Rights of the World Trade Organization) Agreement, the EC found out that there are still major disparities as regards the means of enforcing intellectual property rights. In some Member States, there are no measures, procedures and remedies such as the right of information and the recall, at the infringer's expense, of the infringing goods placed on the market.

Given the fact that

- "the disparities between the systems of the Member States as regards the means of enforcing intellectual property rights are prejudicial to the proper functioning of the Internal Market and make it impossible to ensure that intellectual property rights enjoy an equivalent level of protection throughout the Community…
- the current disparities also lead to a weakening of the substantive law on intellectual property and to a fragmentation of the internal market in this field…
- infringements of intellectual property rights appear to be increasingly linked to organised crime" (IPR Enforcement Directive, Preamble recital 8 et seq.),

effective enforcement of the substantive law on intellectual property should be ensured by specific action at Community level. The objective of this Directive is to approximate legislative systems so as to ensure a high, equivalent and homogeneous level of protection in the internal market.

Thus the general obligations of this Directive (Art. 3) state:

"1. Member States shall provide for the measures, procedures and remedies necessary to ensure the enforcement of the intellectual property rights covered by this Directive. Those measures, procedures and remedies shall be fair and equitable and shall not be unnecessarily complicated or costly, or entail unreasonable time-limits or unwarranted delays.

2. Those measures, procedures and remedies shall also be effective, proportionate and dissuasive and shall be applied in such a manner as to avoid the creation of barriers to legitimate trade and to provide for safeguards against their abuse."

3.3 Code of Good Practice for Copyright in Terminology by Infoterm

The Code of Good Practice for Copyright in Terminology has been developed in the 1990s by Infoterm in cooperation with the legal departments of international organizations. It aims at defining rules of conduct while exchanging, obtaining, and using terminological data.

As terminology work is very labour-intensive and time-consuming, a cooperation between institutions and organizations active in the production of terminological data seems advisable. On the other hand, for the same reasons, high-quality terminology is valuable and should be respected as such. For this purpose the Code of Good Practice defines general ethical provisions regarding the respect of copyright, reference procedures, protection of data integrity, and quoting rules. As long as these issues are not or cannot be put into clear legal provisions or covered by bilateral agreements, the parties should accept the Code as the minimum set of rules, which however are only morally, not legally binding.

Infoterm used the following general observations in developing good practice for copyright in terminology.

Terminological data (terminologies) are important in a number of basic scientific and technical areas, such as

- Domain (specialized) communication;
- Technical writing, translation, localization, internationalization, and related applications;
- Subject field-specific education and training;
- Recording, indexing and retrieval of specialist information, etc.

As a rule, high-quality, reliable terminological data are prepared by teams of experts (e.g. working groups or sub-committees attached to learned societies, scientific and technical associations, research institutions, or terminology standardization bodies). Such preparation of terminological data in the areas of science and technology aims at unifying terminological usage in order to achieve clarity and consistency. In the social sciences and humanities, on the other hand, terminology work is more likely to aim at making conceptual differences transparent.

Terminology work – and especially terminology standardization – is very labour-intensive and time-consuming. Cooperation between institutions and organizations active in the preparation of terminological data should, therefore, be encouraged as much as possible. Exchanging terminological data helps prevent duplication of effort and create consistent terminologies across national, linguistic and subject field boundaries.

Cooperation in terminology preparation, and the exchange of terminological data in particular, may entail:

- Taking over a greater or smaller number of terminological entries or subsets of data from one or more existing terminological entries;
- Exchanging terminological data for use as raw material for systematic terminology work;
- Merging terminological data from different sources to prepare new entries, records, etc.

These activities should take place within the context of the requirements of copyright laws and other laws concerning intellectual property. They should aim both to avoid unduly impeding the exchange of ideas and to give due acknowledgement of the intellectual property of the originator of the data.

While concepts, as “units of knowledge”, should be regarded as the intellectual property of all mankind, their representations as terms and definitions (or other kinds of concept description), or as graphic symbols (or other kinds of non-linguistic representation) must be considered to be the intellectual property of the originator (i.e. a single expert, group of experts, or institution/organization), if this information has been conceived or prepared by the respective originator in the form of a terminological entry, a specific sub-section of an entry, or a collection of terminological data.

All institutions/organizations which prepare terminologies or which own terminological data should regard these as an important contribution to the intellectual property of mankind and should make them available to outside users on terms and conditions which reflect the nature of the terminologies in each case.

Code of Good Practice

Where no bilateral agreements have been concluded to the contrary, the following general provisions shall apply as a code of good practice when importing, entering, or exchanging terminological data:

1. Originators' intellectual property

1.1. Reference to the origin of terminological data shall be explicitly made whenever (all or subsets of) the data are reproduced (output) or passed on to third parties. This applies equally to individual items and to subsets of data from terminological entries or records.

1.2. Where the origin of large volumes of data is to be documented, a single reference to the source may be all that is required when the data are reproduced or transferred. In this case, however, the provider must ensure that the recipient of the data agrees to give due acknowledgement to the originator of the data in all cases.

1.3. Where terminological data have been obtained from an originator who also markets the data himself or herself, the originator's consent shall be obtained where the data exchanged or taken over are made available to a third party in the form of complete entries or as parts of entries.

1.4. Data under copyright must not be passed on without the agreement of the originator. This does not refer to individual entries or a limited set of individual entries which are to be used for research or teaching purposes under the conditions of exemptions from copyright stipulations as they exist in the Berne Convention and its implementations at national level.

1.5. Financial agreements on licenses and royalties must be observed.

1.6. Institutions and organizations, in which large numbers of users have access to terminological data from an external source (i.e. the author{s} themselves or the economic rights holder{s}, such as a publishing house), are responsible for taking all necessary measures against uncontrolled downloading/copying which violates any rights claimed by the originator{s} or rights holders.

2. Data integrity

2.1. Measures to protect data integrity must be strictly observed and must not be deliberately violated (e. g. by introducing minor changes or by taking data out of context). However, the correction of typographic errors and other obvious mistakes is permissible where justified.

2.2. In the case of highly sensitive terminological data (e. g. where safety issues are involved) the strict observance of data integrity with respect both to individual items of information and to data structures shall be obligatory.

2.3. Data marked as secret or confidential must not be passed on without the prior (preferably written) consent of their respective owner.

3. Standardized terminology

3.1. The exchange of terminological data among standards bodies and between standards bodies and relevant specialist institutions and organizations, in order to increase the volume and to improve the quality of standardized terminology, should be encouraged. Given the highly authoritative character of standardised or unified terminologies on the one hand and the highly labour-intensive efforts to create them, cooperation among standards bodies and between them and authoritatively unifying terminologies should be developed as much as possible.

3.2. In the case of terminological records, where no other agreement to the contrary has been made the originating standards body shall be indicated in every individual item or set of terminological information taken over. In this connection national standards bodies should follow the rules, established by international and European organizations of standardization, which regulate observance of copyright when international standards are adopted as regional or national standards.

3.3. Standards bodies should promote active cooperation in terminological data by assigning authoritative foreign language equivalents (and–if possible–definitions as well) to the entries received from sister organizations. Such cooperation needs written bilateral agreements, if federated agreements are not available, stipulating the conditions for the exchange and re-use of the data in accordance with existing legal frameworks.

3.4. Standards bodies and other institutions/organizations considered as authorities in their subject field, are encouraged to collaborate in the harmonization of existing terminologies.

3.5. Cooperation concerning standardized terminologies shall conform to the Code of Good Practice for the preparation, Adoption and Application of Standards, which has to be observed according to the annex of WTO Agreement on Technical Barriers to Trade.

4. Limited quotations of terminological data for scientific, research, teaching and training purposes

As a rule, copyright provisions do not apply

- in cases involving limited extracts of individual terminological data within the limits of defined exemptions from copyright

and

- to the use of individual items of terminological data or entries in scientific publications (limited quotations, fair use etc.) and for teaching and training purposes, provided that no data integrity rules are violated and that correct citation is ensured wherever possible and applicable.

CHAPTER 4

EVALUATION AND DESCRIPTION OF TERMINOLOGY RESOURCES IN EUROTERMBANK

One of the major tasks of the EuroTermBank project was identification, description, and classification of a large number of existing printed and electronic terminology resources available in participating countries and selection of resources for possible inclusion in the EuroTermBank database. In this section approaches for evaluation, selection and description of resources are described.

4.1 Evaluation and selection of resources

In order to evaluate the terminology resources systematically and select and prioritize them for inclusion, several criteria have been used that are described in this section.

4.1.1 Considerations for evaluation of terminology resources

Project partners agreed to deal with Language for Special Purposes (LSP) only and exclude the Language for General Purposes (LGP) resources. The project is dealing with terminology, defined in ISO 1087-1:2000 as a "set of designations... belonging to one special language", special language being defined as "language used in a subject field and characterized by the use of specific linguistic means of expression".

The institutions or the authors creating terminology resources can be considered a valuable indication of the quality of a collection. When the institution or the author is known for well-founded terminology work and reputed exerts of the respective subject field are involved, there are good chances that the quality of the terminology collection is appropriate. However, just the fact that an institution or an author is not known so far should not be a sufficient reason to exclude their terminology resources from consideration.

Data originators listed by degree of authoritativeness are:

- Legal international or national authority determined by legislation or jurisdiction
- Officially authorized harmonization/standardization body
- Institution authorized or recognized as a subject field authority
- Formally or informally recognized subject-field authority
- Non-authoritative terminology source

Another important criterion is the methodological approach the terminology resource is based on – whether the relevant national or international standards for terminology work have been observed. Central quality criteria are concept orientation, systematic choice of concepts, subject field indication and usage of notes, alphabetical indices in all languages, abbreviations, definitions, grammatical information, phonetic information, target group mentioned etc.

Access conditions are also important for the creation of a publicly accessible terminology bank. To make use of the data, either the terminology resources must be freely accessible or the respective copyright holder should be ready to cooperate and conclude a copyright agreement with the project consortium.

Actuality of data is another critical criterion for the selection of terminology resources. This criterion is closely connected with the respective subject field and the purpose the terminology collection has been created for.

4.1.2 Guidelines and criteria for the evaluation of terminology resources

The following table presents guidelines and criteria applied when selecting resources for the EuroTermBank project:

Criteria	Guidelines and descriptions
General criteria	
What is a resource	▪ monolingual terminology (covering one or more subject-fields) ▪ multilingual terminology (covering one or more subject-fields) How many records for a given subject-field constitute a resource? What is the minimum number of records in a given language (or multilingual) to be considered as one resource (provided that the data of the language(s) can be considered as complete)?
What is the 'value' of terminological data	▪ degree of authoritativeness of the data originator ▪ quality of data documentation used and references hereto (viz. verifiability) ▪ preparation by ▪ group of experts ▪ one or few experts ▪ specialized lexicographers ▪ others ▪ 'completeness' of data (which may vary according to different conventions in different subject-fields) ▪ 'up-to-datedness' of data (date of input/latest revision should be quite recent in highly dynamic fields) ▪ existence of a (internal/external) validation mechanism
Vertical evaluation criteria (by subject-field)	
Authoritative nature of data (degree of authoritativeness)	▪ according to the status of the data originator being ▪ a legal or quasi-legal (public or semi-public) authority ▪ a harmonizing/standardizing (or quasi-standardizing) body ▪ an 'informal' authority in the respective subject-field As a rule there is no absolute 'authority' covering all applications, the authority in most cases is restricted to a (implicitly or explicitly) defined scope, but can often be extended towards similar/neighboring applications.
Legal (or quasi-legal)	▪ determined by legislation or jurisdiction at international, European or national levels

Criteria	Guidelines and descriptions
Harmonized/ standardized	▪ by an official public or officially authorized harmonization/ standardization body
Quasi-standardized	▪ by a subject-field authority recognized (e.g. IUPAC) or by an institution/organization authorized for this purpose, but not belonging to the official standardization framework, e.g. technical rules issued by public administration: ▪ prepared within the framework of a working group or committee/ commission established for this purpose ▪ prepared on the basis of a contract/mandate given to one (or more) expert(s)
Issued by a (formally or informally recognized) subject-field authority	▪ prepared within the framework of a working group or committee/ commission established for this purpose ▪ prepared by one (or more) individual experts on behalf of the subject-field authority ▪ adopted by the subject-field authority from outside originators and a. prepared on the basis of a proper terminological methodology (e.g. following the respective ISO standards) ▪ individual data being well documented ▪ (incl. indication of source references, originating body/expert etc., responsibility codes etc.) ▪ prepared by (individual or a group of) subject-field experts ▪ prepared by other kind of expert(s) (e.g. specialized lexicographer, translator, etc.)
Non-authoritative terminology	▪ prepared within the framework of a working group or committee/ commission established for this purpose ▪ prepared by one (or more) individual experts on behalf of an issuing institution/organization (e.g. publisher) ▪ adopted by an issuing institution/organization from outside originators and ▪ prepared on the basis of a proper terminological methodology (e.g. following the respective ISO standards) ▪ individual data being well documented (incl. indication of source references, originating body/expert etc., responsibility codes etc.) ▪ prepared by (individual or a group of) subject-field experts ▪ prepared by other kind of expert(s) (e.g. specialized lexicographer, translator, etc.)
Horizontal evaluation criteria (common to all subject fields)	
High quality of documentation of data	▪ facilitating the verifiability of data
High degree of detail and completeness	▪ leading to clarity/transparency of data structure ▪ resulting in multifunctional terminological data The above-mentioned principle criteria do not preclude the possibility of high-quality data prepared by non-authoritative originators in individual cases. They are to a large extent similar to the ('formal') quality criteria in QA.

Criteria	Guidelines and descriptions
Degree of authoritativeness in relation to costs of preparation	Typically, terminological data prepared ▪ by groups or teams in an authoritative framework tend to be costlier than those prepared by one or few individuals; ▪ in a highly systematic and well documented way tend to be costlier than those prepared in an unsystematic way; ▪ by experts tend to be costlier than those prepared by non-expert terminographers. The costs for preparing terminological data may vary from USD 10 (by a non-expert terminographer in a well-documented and comparatively less dynamic subject field) per entry in a given language to x1000 USD (by highly authoritative expert groups) per entry – if all costs are calculated. Costs for the preparation of terminologies are often disconnected from the price. Generally, the ‚price' of terminological data is far below their ‚creation costs'.

4.2 Terminology resource description (TeDIF)

One of the major project tasks was to identify and describe terminology resources in new EU member countries. Due to a large number of resources to be described and different organizations in several countries involved in this process it was important to use a common format for resource description. For this purpose the TeDIF format has been chosen.

The Terminology Documentation Interchange Format TeDIF was developed in the framework of the TDCnet project – European Terminology Documentation Centre Network, co-funded by the EU Commission. The TeDIF format was developed with the purpose to establish a common format for bibliographical and factual data related to terminology. These include in detail:

1. Bibliographical data

- literature (serials, monographs, articles, journals, theses, etc.)
- term collections (printed dictionaries, glossaries, thesauri, classifications, terminology databases, etc.)

2. Factual data

- corporate entities (organizations, institutions)
- persons (experts)
- projects
- terminology management software
- events (conferences, workshops)
- teaching and training opportunities

For the purpose of the EuroTermBank project, TeDIF was slightly adapted. TeDIF information types were limited to the description of term collections (full TeDIF specification also allows descriptions of other bibliographical and factual data like projects, events and persons that are not required in EuroTermBank). Other modifications included a possibility to multiply the fields describing the author and copyright holder according to the number of persons/organizations and the addition of fields for the indication of the languages of definitions and context information.

TeDIF is an SGML-based format (Standard Generalized markup Language, ISO 8879:1986) to describe and exchange data. Since TeDIF is also XML-compatible (Extended markup Language, subset of SGML), it is open to the newest developments in markup languages, the usage of Unicode, and an easier conversion to HTML and other formats.

Some of the project partners with more advanced technical skills prepared resource description directly in TeDIF format. For other partners a special Excel spreadsheet form was created providing an easy way for entering data and avoiding possible mistakes. In order to validate and transform Excel files to the TeDIF format a converter utility was programmed.

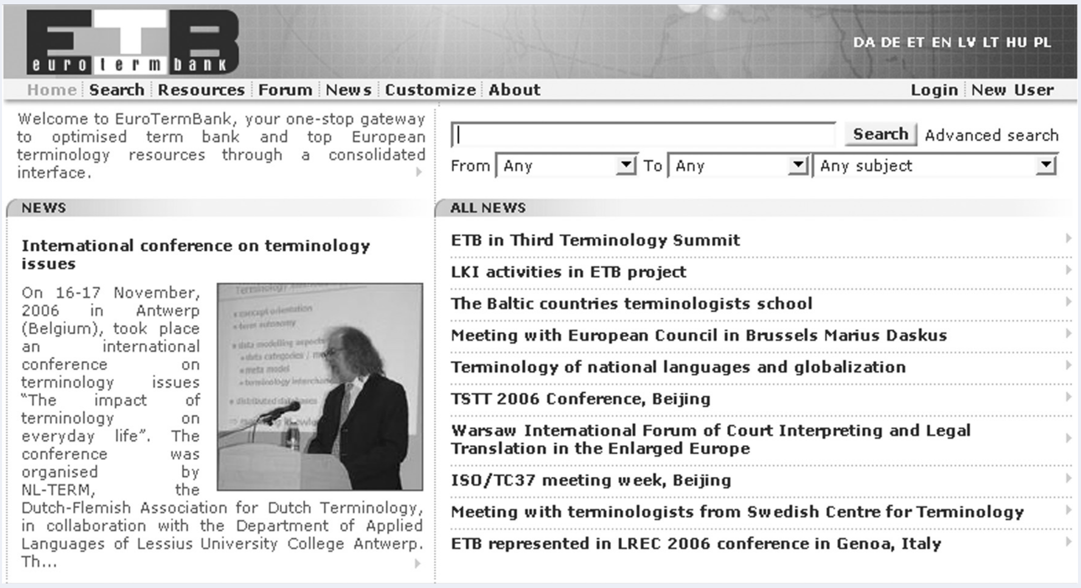
TeDIF continues to be used for importing terminology resource meta-data into the EuroTermBank database, as well as for the consolidation and analysis of data.

CHAPTER 5

THE EUROTERMBANK PORTAL

This chapter provides an overview of the multilingual EuroTermBank (ETB) portal, the most tangible of the project resources, available at www.eurotermbank.com. It gives an overview of the portal and its services from the user point of view, as well as describes its architecture and data structure.

The following illustration shows the homepage of the ETB portal:



Welcome to EuroTermBank, your one-stop gateway to optimised term bank and top European terminology resources through a consolidated interface.

International conference on terminology issues

On 16-17 November, 2006 in Antwerp (Belgium), took place an international conference on terminology issues "The impact of terminology on everyday life". The conference was organised by NL-TERM, the Dutch-Flemish Association for Dutch Terminology, in collaboration with the Department of Applied Languages of Lessius University College Antwerp. Th...

ETB in Third Terminology Summit

LKI activities in ETB project

The Baltic countries terminologists school

Meeting with European Council in Brussels Marius Daskus

Terminology of national languages and globalization

TSTT 2006 Conference, Beijing

Warsaw International Forum of Court Interpreting and Legal Translation in the Enlarged Europe

ISO/TC37 meeting week, Beijing

Meeting with terminologists from Swedish Centre for Terminology

ETB represented in LREC 2006 conference in Genoa, Italy

Figure 9: Homepage of the EuroTermBank portal.

5.1 System overview

The EuroTermBank system is an integrated termbank service, providing a unified access to multiple terminology resources, and an interface to publishing terminology. It provides:

- A single access point to all terminology needs of a user providing continuously extended and updated content by adding new terminology resources and adding new data;
- Access to terminology – query schemes suitable for particular usage scenarios;
- Publishing terminology – a service for terminology authors/providers to provide input to the system.

The EuroTermBank system is based on open data exchange standards. The system is accessible primarily through a Web browser. The users can pick a system interface language depending on their preferences.

The system performs user authorization, distinguishing between the following user groups: Anonymous Users, Registered Subscribers, Editors, and Administrators.

Depending on their user role, these users can authenticate themselves, search for terminology, participate in discussions and give feedback to the content and system developers, edit terminology entries, or administer the system.

Administrators can add or import new terminology collections, export a subset of the data in the system, delete terminology collections, create, edit and delete users and perform other administrative tasks.

If a user requests a term in a language or an industry sector not present in the database, search for the requested information in external sources is available. This feature is an innovative approach in terminology databases, as most other terminology databases allow searching only in resources stored in the particular database. EuroTermBank system supports querying external databases and merges the results from many sources in one search result list. External databases usually have very different formats, therefore such external database querying and result harmonization and unification is an important part of the project; it significantly influences data structure, data categories, exchange formats and system architecture described in this chapter.

The system logs its critical activities allowing basic audit and reporting.

5.2 Usage scenarios

To understand the potential EuroTermBank user needs, a survey of different groups of potential system users was carried out. The typical usage scenarios of terminological resources were the following (in decreasing order of popularity):

- Translation – Most users look to a terminological service for translational terminology. They require integration of multiple data sources and convenient user interface. It may also be a requirement to provide integration with popular CAT (computer-assisted translation) tools.
- General research („look up terms") – This is a vague concept, but the action itself usually happens during research and reading. This requires comprehension and research assistance, which a terminological service must address.
- Lexicography – The terminology system is a research tool when used by lexicographers, mostly through the stored definitions. For this purpose, integration of multiple data sources is essential.
- Terminology manipulation – services for those who are (a) building their own termbases – this is basically a research facility with advanced filtering and export features, (b) providing input to the terminology bank.
- Adding or changing entries – This is an interface of the service for terminologists providing input, i.e. integrating or aiming to integrate their terminology resources with the terminology bank. For their purposes, a standardized interface is required where (a) external terminology resources can be plugged into the service, (b) terminology resources are uploaded to system servers, (c) existing terminology resources are manipulated.

5.3 EuroTermBank services

This section describes the main features of the EuroTermBank portal. The specific functions of ETB can be divided into several groups: user authentication, search, editing, administration, feedback and communication with external databases, data import and export.

5.3.1 Search

EuroTermBank system is an integrated platform accessible online with most popular Internet browsers. EuroTermBank supports hyperlinks and pictures in the binary data fields.

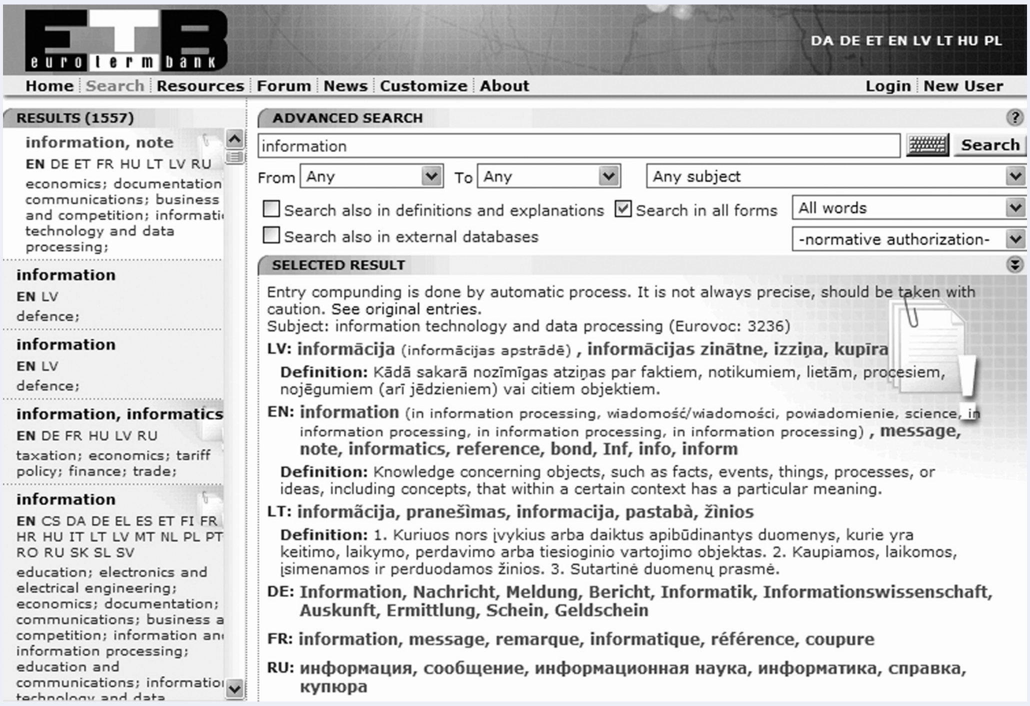
One of the main features for users of the ETB system is search within the database. User interface enables the users to query the full database containing terms from all subjects, or choose one particular subject to search. They are able also to search in any language presented in the database and even search terms in all languages. Additional important search feature provided by user interface is that if the user is requesting a term in a language or a domain that are not present in the database, this request can be forwarded to another external database.

Communication functions allow search of necessary information in external databases and ensure connection and disconnection.

Search function provides simple and full text search.

It is also possible to search and display entries from external databases. The ext_db table containing resource names and descriptions is used to list all available external resources.

The following illustration shows the search results page of the ETB:



RESULTS (1557)

information, note
EN DE ET FR HU LT LV RU
economics; documentation communications; business and competition; informati technology and data processing;

information
EN LV
defence;

information
EN LV
defence;

information, informatics
EN DE FR HU LV RU
taxation; economics; tariff policy; finance; trade;

information
EN CS DA DE EL ES ET FI FR HR HU IT LT LV MT NL PL PT RO RU SK SL SV
education; electronics and electrical engineering; economics; documentation; communications; business a competition; information an information processing; education and communications; informatio technology and data

ADVANCED SEARCH

information | Search

From Any | To Any | Any subject

□ Search also in definitions and explanations ☑ Search in all forms | All words

□ Search also in external databases | -normative authorization-

SELECTED RESULT

Entry compunding is done by automatic process. It is not always precise, should be taken with caution. See original entries.
Subject: information technology and data processing (Eurovoc: 3236)

LV: informācija (informācijas apstrādē) **, informācijas zinātne, izziņa, kupīra**

Definition: Kādā sakarā nozīmīgas atziņas par faktiem, notikumiem, lietām, procesiem, nojēgumiem (arī jēdzieniem) vai citiem objektiem.

EN: information (in information processing, wiadomość/wiadomości, powiadomienie, science, in information processing, in information processing, in information processing) **, message, note, informatics, reference, bond, Inf, info, inform**

Definition: Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning.

LT: informãcija, pranešìmas, informacija, pastabà, žìnios

Definition: 1. Kuriuos nors įvykius arba daiktus apibūdinantys duomenys, kurie yra keitimo, laikymo, perdavimo arba tiesioginio vartojimo objektas. 2. Kaupiamos, laikomos, įsimenamos ir perduodamos žinios. 3. Sutartinė duomenų prasmė.

DE: Information, Nachricht, Meldung, Bericht, Informatik, Informationswissenschaft, Auskunft, Ermittlung, Schein, Geldschein

FR: information, message, remarque, informatique, référence, coupure

RU: информация, сообщение, информационная наука, информатика, справка, купюра

Figure 10: Example of a results page in the EuroTermBank portal.

5.3.2 Editing

Editing allows the creation of new entries and editing of the existing ones as well as viewing the editing history of the entry. This service is accessible to the Editors user group.

Users having the Editor role have the rights to edit terminology data explicitly specified for them. To specify the data subset that a particular user can edit, terminology collections are defined. These collections are stored in table collections. Each collection can have a number of terms. The may_edit table containing collection and user identifiers is used to specify which collections a user may edit.

5.3.3 Import

The authors and owners of terminology collections can provide their sources in the form of a text file, database, pdf file or even as a printed publication. EuroTermBank specialists process the material to transform it into the XML structure that is used to store and manage the terminology and related information in the internal database.

All terminology collections currently in the ETB database have gone through the import process, preceded by a certain amount of processing and preparation.

5.3.4 Export

The ETB system provides the ability of exporting a subset of terminology entries stored in the internal database as TBX (the native format of the database) or as formatted text. After specifying the required subset parameters, this is an automated process.

Only portal administrators are allowed to perform export due to copyright, security and business reasons.

5.3.5 Access for 3rd party software

Manufacturers of CAT tools and similar software can provide their users direct access to the contents of the ETB internal database by the help of customized APIs. This feature is especially useful in the translation scenario.

5.3.6 EuroTermBank discussions services

As a portal, EuroTermBank provides public discussions services, enabling various users – translators, subject matter experts, lexicographers, terminologists, and others, to exchange ideas and provide comments. ETB also contains various public notices and documents in read-only mode.

5.4 System architecture

5.4.1 Overview

The EuroTermBank (ETB) system has multi-tier architecture as illustrated below, with separate user interface, business logic and database layers.

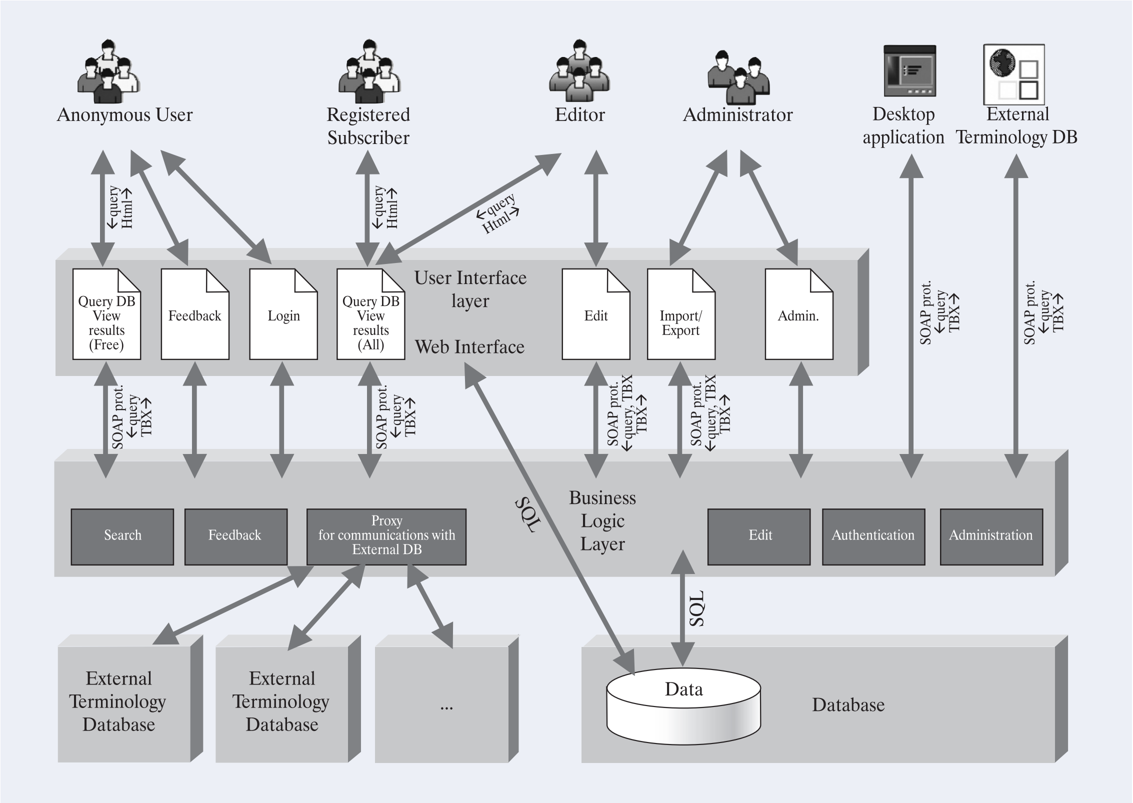


Figure 11: Architecture of the EuroTermBank system.

Users can access the system in the following ways:

- Normally users access the system through a web browser
- Users are able to access the system through desktop applications developed by Independent Software Vendors (ISVs)
- Users are able to access the system through other term banks which support query and data exchange with the EuroTermBank system.

5.4.2 Types of users

The EuroTermBank system has several user groups. By default, users access the system through web interface and get access permissions of the Anonymous User group. The following table shows the privileges of each user group:

Types of users	Privileges
Anonymous user	▪ query the free part of resources in the system and view the results ▪ provide feedback about the system and terminology in general ▪ log into the system to get access rights of Registered Subscriber, Editor or Administrator
Registered subscriber	▪ query resources in the system and view the results (free and fee-based, depending on type of subscription) ▪ participate in discussions
Editor	▪ perform all the actions available to the users of the Registered Subscriber group ▪ access parts of the terminology database for editing Each user of this group is able to edit only the part of the database explicitly granted to this user.
Administrator	▪ manage system user accounts ▪ perform system maintenance tasks, such as making backups, importing of new terminology databases, etc. ▪ perform special tasks for customers who need special fee-based services, e.g. exporting or printing some parts of database in special format ▪ access the business logic functions and database directly without user interface

When external desktop applications or external terminology databases access the system they have to identify themselves. The system checks the kind of external agent who is accessing it and the user group to which it belongs. Depending on the particular agreement with the EuroTermBank Consortium, external desktop applications or external terminology databases are able to access the system with permissions of Anonymous Users or Registered Subscribers. The external desktop applications or external terminology databases have to send not only data query but also their login information with username and password.

5.4.3 User interface layer

Users are able to access the system through a web browser or through applications developed by ISVs, web browser being the main access channel. The user interface layer in the System Architecture diagram above outlines the user interface elements.

User interface layer of the system does not describe desktop applications developed by ISVs. Although they have access to the system they are not a part of it. These applications access services in the business logic layer directly.

The main user interface elements (or web pages) are the following:

- Search page allowing users to enter queries and browse results
- Forum page allowing users to enter feedback
- Login page allowing users to log into the system (user authentication)

- Editor page allowing database editing
- Resources page that contains general information about resources included in the EuroTermBank system
- News page that contains news and general information related to the EuroTermBank system and terminology
- Administrators pages

The user interface is developed using ASP.NET technology.

5.4.4 Business logic layer

The user interface layer performs only user interface related tasks, such as receiving user input and forming of web pages, while all actual work is done in the business logic layer. This layer performs several different tasks, implemented as separate system modules – search and result formation, feedback, proxy for communications with external databases, database editing, user authentication and administration.

Direct connection to the database from the user interface layer is performed only in the case of irrelevant and unprotected data transportation due to transfer speed considerations.

All business logic functions are implemented as web services using the SOAP standard communication protocol.

5.4.5 Database

The ETB system stores all data in an SQL database. There are several types of data stored in the database – terminology data, user information, feedback, interface elements, and information about external databases. Terminological data are stored in the EuroTermBank data exchange format (TBX) as XML. Part of the information is stored and indexed for search optimization purposes.

5.4.6 External terminology databases

The EuroTermBank system is designed to access external terminology databases through web interface provided by owners of these external databases.

The communication between the business logic layer and external terminology databases goes through internet. External terminology databases can be very different and each may have different protocol and data exchange format. Therefore, a special proxy for communication with external databases to query and exchange data is developed for each external database. This proxy has a special interface for each particular external database connection module.

5.5 Data structure and exchange format

This section provides information and the data structure and data categories, as well as the exchange format and mechanisms used in the EuroTermBank project.

5.5.1 Data structure

The data structure developed for EuroTermBank comprises up to 4 hierarchical levels based on ISO standards 12200 and 12620:

- The entry level provides concept-related data categories applying to all languages. It contains language-independent information like entry identifier, subject information, data collection, administrative information like subset owner identifying institution responsible for the entry, originator, origination date, updater, modification date and a number of other fields.
- The language level provides concept-related data categories applying to the specific language. It contains language-specific information like definition, reference, explanation and other, as well as administrative information.
- The term level provides term-related data categories applying to the specific term. It includes term-related information like term in particular language, entry source, search term containing related forms of the term to facilitate searching, reference with source(s) of the term, usage information and other.
- The word level provides word-related data categories applying to the specific words of a term. As a term may be a multiword string, this level is created to contain lexical information that concerns the individual words of a term. Data categories for lexical information are, for example, part of speech, grammatical number, grammatical gender etc.

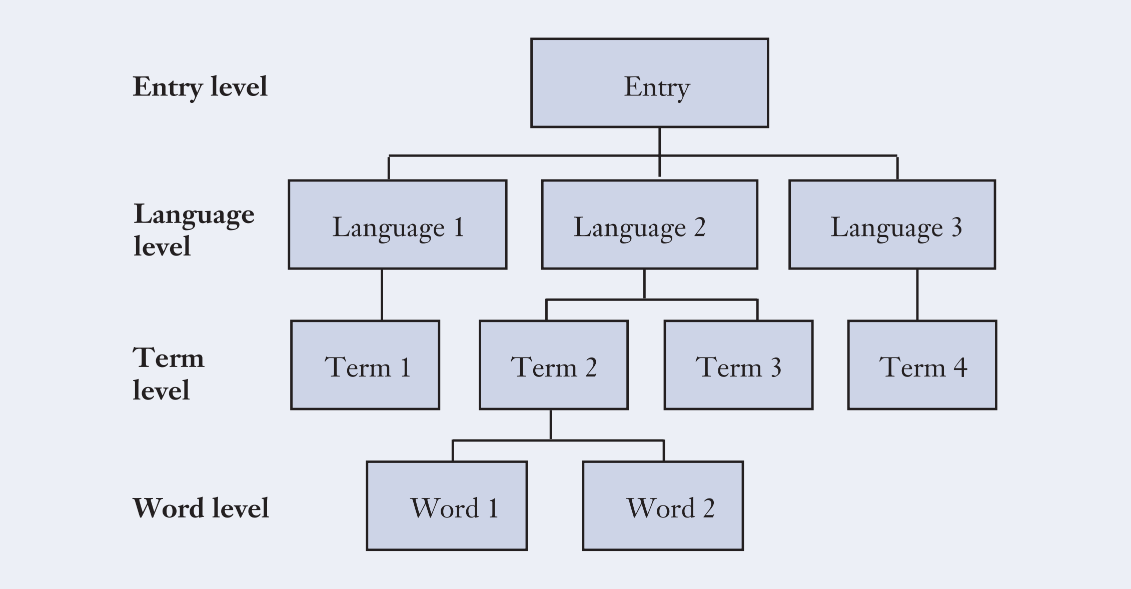


Figure 12: Data structure in the EuroTermBank portal.

5.5.2 Data categories

It is essential that the data structure be based on standards to ensure exchangeability with other data collections and to ensure that data categories are recognizable for outside users. Terminology data structure should comply with ISO standards 12200 and 12620. The original ISO 12620 was designed specifically for concept-oriented terminology management systems but it is also targeted for a broader usage in different terminology applications.

The EuroTermBank data structure comprises information about the concept, the terms that designate the concept and the words that constitute the individual terms. As a multilingual system, it permits definitions in all languages, therefore conceptual information is grouped in two levels: the entry level containing language independent information and the language level containing language specific information. Term related information is included at term level; an example of an information type that might appear at term level is usage information. Lexical information concerning a specific word is included at word level.

The table below describes the data categories for each level of an entry. 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. Although these data categories comply with ISO 12620, this is by no means an exhaustive list of ISO 12620 data categories; the standard contains multiple possibilities that must be considered in relation to the specific application.

Levels	Descriptions
Entry level	
Entry identifier	The value of this data category is a system-generated number that will identify the entry uniquely.
Subset owner	The value of this data category is the institution responsible for the whole entry. As the data collection will contain contributions from many different organizations it is necessary to state clearly who is responsible for the mainteance of each entry.
Originator	An identifier of the person who prepared the entry.
Inputter	An identifier of the person who types in the information.
Origination date	The date the entry was first created.
Updater	The value of this field is the person having made the latest changes to the information at entry level.
Modification date	The date when the latest changes to the entry level were made.
Security subset	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	The data category(ies) chosen for subject information will contain the domain of the particular concept.
Note	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	Contains, for example, tables, figures, videos and other binary data.

Levels	Descriptions
Reference	Reference(s) to the non-textual information.
Data collection	This field can be used to signify that a particular concept belongs in a particular collection of concepts.
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	A reference to other concepts in various ways related semantically to the concept in question, for example broader concept, subordinate concept or related concept.
Language level	This level can contain the following administrative entry-level fields – Originator, Inputter, Origination date, Updater, Modification date (see descriptions above).
Language symbol	Contains the language symbol of the particular language. The symbols specified in ISO 639 should be used.
Non-textual information and Reference	See comment about non-textual information at entry level.
Definition	A formal and precise description of the concept.
Reference	Reference(s) to where the definition given above was found.
Explanation	Compared to the *Definition* field, this field makes it possible to give a more informal description of the concept. This field is particularly useful in cases where a formal definition has not been obtainable.
Note	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 code	Reliability codes are suggested at language and term levels. A reliability code at the language level provides an assessment of the correctness and precision of the information given in relation to the specific concept.
Term level	Originator, Imputer, Origination date, Updater, Modification date – Contains the same information as in levels above.
Entry source	If the entry is imported from another resource, this field contains information about the database or format from which data are imported.
Search term	Contains 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*.
Term	Contains the term: a designation of a defined concept in a specific language by a linguistic expression.
Term type	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.

Levels	Descriptions
Reference	Source(s) of the *term*.
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	A general comment that applies to the entire term level.
Reliability code	Reliability codes are suggested at language and term levels. A reliability code at the term level provides an assessment of the correctness and precision of the information given in relation to the specific term.
Validation information	Validation information is located at term level and not at the other levels though a validation procedure includes validation of all levels.
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. This level can contain the following administrative fields – Originator, Inputter, Origination date, Updater, Modification date.
Term element	Concerns a particular word that forms part of a term.

5.5.3 Data exchange format

Data exchange mechanisms are developed for the EuroTermBank project to enable term import, export and exchange with other terminology databases. The data exchange format is based on TBX (TermBase exchange) format: an open XML-based standard for terminological data exchange developed by LISA, the Localization Industry Standards Association. TBX complies with the terminology markup framework defined by ISO 16642; it specifies a set of data categories from ISO 12620 and adopts an XML style compatible with ISO 12200.

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.

The EuroTermBank system implements the TBX standard with required data categories to enable:

- data exchange between different ETB modules;
- data exchange between external terminology databases;
- data import and export to and from the ETB terminology database;
- data store in the ETB terminology database;
- data editing.

For a more detailed description of the exchange format, see Chapter 1, Methodology recommendations in terminology management, Exchange format. For a detailed specification of the exchange format, see Appendix B (The ETB data exchange format).

Although TBX serves as the universal data exchange format, in practice EuroTermBank deals with terminology content from different institutions that typically store data in a large variety of formats – plain text files, text editor documents, spreadsheets, different types of databases, etc. All the necessary information is collected from these files and the missing information is added, considering the obligatory fields listed in the data structure requirements.

To import terminology data into the ETB database, data must be structured according to the EuroTermBank TBX-compliant data exchange format. To convert terminology resources to this specific format, a number of conversion tools is required. As each resource is structured differently, an individual converter is developed or adapted for each resource type, as it is impossible to create one universal converter. Most of the tools to convert data to the TBX format are written in the Perl computer language, as it provides a powerful regular expression engine built directly into its syntax and open source support is available.

APPENDICES

Appendix A. Data structure

Data category	ISO 12620 position code	Description
Entry level		
Administrative information		
Entry identifier	A.10.15	The value of this data category is a system-generated number that will identify the entry uniquely.
Subset owner	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	A.10.02.02.01	An identifier of the person who prepared the entry.
Inputter	A.10.02.02.02	An identifier of the person who types in the information.
Origination date	A.10.02.01.01	The date the entry was first created.
Updater	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	A.10.02.01.03	The date when the latest changes to the entry level were made.
Security subset	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	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.

Data category	ISO 12620 position code	Description
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.
Language level		
Administrative information		
Originator	A.10.02.02.01	An identifier of the person who prepared the language level.
Inputter	A.10.02.02.02	An identifier of the person who types in the information.
Origination date	A.10.02.01.01	The date the language level was first created.
Updater	A.10.02.02.03	An identifier of the person having made the latest changes to the information at language level.
Modification date	A.10.02.01.03	The date when the latest changes to the language level were made.
Language symbol	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	A.05.01	In this field, a formal and precise description of the concept is given.

Data category	ISO 12620 position code	Description
Reference		Reference(s) to where the definition given above was found.
Explanation	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	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	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.
Term level		
Administrative information		
Originator	A.10.02.02.01	An identifier of the person who prepared the term level.
Inputter	A.10.02.02.02	An identifier of the person who types in the information if this person varies from the originator.
Origination date	A.10.02.01.01	The date the term level was first created.
Updater	A.10.02.02.03	An identifier of the person having made the latest changes to the information at term level.
Modification date	A.10.02.01.03	The date when the latest changes to the term level were made
Entry source	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	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

Data category	ISO 12620 position code	Description
Terms		
Term	A.01	This field will contain the term: a designation of a defined concept in a specific language by a linguistic expression.
Term Type	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	A.08	A general comment that applies to the entire term level.
Reliability		
Reliability 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.

Data category	ISO 12620 position code	Description
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	A.10.02.02.01	An identifier of the person who prepared the word level.
Inputter	A.10.02.02.02	An identifier of the person who types in the information.
Origination date	A.10.02.01.01	The date the entry was first created.
Updater	A.10.02.02.03	An identifier of the person having made the latest changes to the information at word level.
Modification date	A.10.02.01.03	The date when the latest changes to the word level were made.
Word		
Term element	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	A.02.05	This data category contains a representation of the pronunciation of a word.

Appendix B. The ETB data exchange format

Tag & sample	ISO 12620	Required	Description
`<?xml version=1.0 encoding=utf-8?>`			
`<!DOCTYPE martif PUBLIC ISO 12200:1999A//DTD MARTIF core (DXFcdV04)//EN TBXcdv04.dtd>`			
`<martif type=TBX xml:lang=en>`			
`<martifHeader>`			
`<fileDesc>`			
`<titleStmt>`			
`<title>Title of the collection</title>`		◆	Title of the collection
`</titleStmt>`			
`<sourceDesc>`			
`<p>Description of the collection source</p>`			Description of the source
`</sourceDesc>`			
`</fileDesc>`			
`<encodingDesc>`			
`<p type=DCSName>TBXDv04Cycom.xml</p>`			File with encoding description
`</encodingDesc>`			
`</martifHeader>`			
`<text>`			
`<body>`			
`<termEntry id=’ID67’>`	A.10.15	◆	Entry identifier a system-generated number that will identify the entry uniquely
`<admin type=’sourceLanguage’>en</admin`	A.10.23		Source Language the source language of a set of terms that are not perfectly multi-directional
`<admin type=’subsetOwner’>SIA TILDE</admin>`	A.10.02.02.10	◆	Subset owner institution responsible for the whole entry

Tag & sample	ISO 12620	Required	Description
<admin type=’securitySubset’>2</admin>	A.10.03.09	◆	Security subset a security classification expressing the confidentiality level of the entire entry
<transacGrp>		◆	
<transac type=transactionType>origination</transac>		◆	
<transacNote type=’responsibility’>R. Smith</transacNote>	A.10.02.02.01	◆	Originator an identifier of the person who prepared the entry
<date></date>	A.10.02.01.01	◆	Origination date The date the entry was first created
</transacGrp>		◆	
<transacGrp>		◆	
<transac type=transactionType>creation</transac>		◆	
<transacNote type=’responsibility’>J. Smith</transacNote>	A.10.02.02.02	◆	Inputter An identifier of the person who types in the information
</transacGrp>		◆	
<transacGrp>			
<transac type=transactionType>modification</transac>			
<transacNote type=’responsibility’>J. Clarck</transacNote>	A.10.02.02.03		Updater the person having made the latest changes to the information at entry level
<date></date>	A.10.02.01.03		Modification date The date when the latest changes to the entry level were made
</transacGrp>			
<descrip type=’subjectField’>23</descrip>	A.04	◆	Subject Field the subject of the concept

Tag & sample	ISO 12620	Required	Description
<note>more subject information</note>	A.08		Note a note related to the classification number
<descrip type=’otherBynaryData’>235j239sd21</descrip>	A.05.05.05		Other binary data
<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>	A.10.20		Reference
<admin type=’projectSubset’>abc</admin>	A.10.03.03		Project subset an identifier of a particular collection of concepts
<descrip type=’broaderConceptGeneric’ target=’entryId’> </descrip>	A.07.02.01		Broader concept
<descrip type=’subordinateConceptGeneric’ target=’entryId’></descrip>	A.07.02.03		Subordinate concept
<descrip type=’relatedConcept’ target=’entryId’></descrip>	A.07.02.05		Related concept
<langSet lang=en’>	A.10.07	◆	Language symbol the language symbol of the particular language
<transacGrp>		◆	
<transac type=transactionType>origination </transac>		◆	
<transacNote type=’responsibility’>R. Smith</transacNote>	A.10.02.02.01	◆	Originator an identifier of the person who prepared the language level
<date></date>	A.10.02.01.01	◆	Origination date The date the language level was first created
</transacGrp>		◆	
<transacGrp>		◆	
<transac type=transactionType>creation</transac>		◆	
<transacNote type=’responsibility’>J. Smith</transacNote>	A.10.02.02.02	◆	Inputter An identifier of the person who types in the information
</transacGrp>		◆	
<transacGrp>			
<transac type=transactionType>modification</transac>			

Tag & sample	ISO 12620	Required	Description
<transacNote type='responsibility'>J. Clarck</transacNote>	A.10.02.02.03		Updater the person having made the latest changes to the information at language level
<date></date>	A.10.02.01.03		Modification date The date when the latest changes to the language level were made
</transacGrp>			
<descrip type='otherBynaryData'>235j239sd21</descrip>	A.05.05.05		Other binary data
<admin type='sourceIdentifier' target='DIN-561.12'>p.21</ref>	A.10.20		Reference
<note>more inf about the concept in particular language</note>	A.08		Note A note field related to the entire language level
<descrip type='reliabilityCode'>2</descrip>	A.03.04		Reliability code an assessment of the correctness and precision of the information given in relation to the specific concept.
<descripGrp>			
<descrip type='definition'>degree of obstruction</descrip>	A.05.01		Definition
<admin type='sourceIdentifier' target='DIN-561.12'>p.21</ref>	A.10.20		Reference a reference to the definition
</descripGrp>			
<descripGrp>			
<descrip type='explanation'>degree of obstruction</descrip>	A.05.02		Explanation
<admin type='sourceIdentifier' target='DIN-561.12'>p.21</ref>	A.10.20		Reference A reference to the explanation
</descripGrp>			
<ntig>			
<transacGrp>			

Tag & sample	ISO 12620	Required	Description
<transac type=transactionType>origination</transac>			
<transacNote type='responsibility'>R. Smith</transacNote>	A.10.02.02.01	◆	Originator - an identifier of the person who prepared the term level
<date></date>	A.10.02.01.01	◆	Origination date - The date the term level was first created
</transacGrp>			
<transacGrp>			
<transac type=transactionType>creation</transac>			
<transacNote type='responsibility'>J. Smith</transacNote>	A.10.02.02.02	◆	Inputter - An identifier of the person who types in the information
</transacGrp>			
<transacGrp>			
<transac type=transactionType>modification</transac>			
<transacNote type='responsibility'>J. Clarck</transacNote>	A.10.02.02.03		Updater – the person having made the latest changes to the information at term level
<date></date>	A.10.02.01.03		Modification date - The date when the latest changes to the term level were made
</transacGrp>			
<transacGrp>			
<transac type=transactionType>approval</transac>			
<transacNote type='responsibility'>R. Smith</transacNote>	A.10.02.02.04		Approver – An identifier of the person consolidating the entry
<date></date>	A.10.02.01.04		Approval date
</transacGrp>			
<termGrp>			

Tag & sample	ISO 12620	Required	Description
`<admin type=’entrySource’>db</admin>`	A.10.13		Entry source the database or format from which data are imported
`<admin type=’intellectualPropertyRights’ >p.21</admin>`	No ISO Code		Intellectual property rights
`<descrip type=’context’>state transition table</descrip>`	A.05.03		Context
`<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>`	A.10.20		Reference Source(s) of the context example
`<termNote type=’register’ >neutralRegister</termNote>`	A.02.03.03		Register a classification indicating the relative level of language assigned to a term
`<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>`	A.10.20		Reference Reference(s) to the register information
`<termNote type=’temporalQualifier’ >archaicTerm</termNote>`	A.02.03.05		Temporal qualifier Information about a term with respect to its use over time
`<termNote type=’usageNote’ >rarely used</termNote>`	A.02.03.01		Usage note local, regional or geographic usage of the term
`<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>`	A.10.20		Reference Reference(s) to the Usage note field.
`<note>general note to term level</note>`	A.08		Note A general comment that applies to the entire term level
`<descrip type=’reliabilityCode’>4</descrip>`	A.03.04		Reliability code an assessment of the correctness and precision of the information given in relation to the specific term

Tag & sample	ISO 12620	Required	Description
<termNote type=’normativeAuthorization’>preferredTerm</termNote>	A.02.09.01		Normative authorization A term status qualifier assigned by an authoritative body
<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>	A.10.20		Reference Reference to the normative organization
<admin type=’searchTerm’>transition table</admin>	A.10.06.03		Search term related forms of the term to facilitate searching
<term>transition table</term>	A.01	◆	Term
<termNote type=’termType’ >fullForm</termNote>	A.02.01	◆	Term Type Some possible values are: main entry term, abbreviation, acronym, short form, variant, formula, synonym ….
<admin type=’sourceIdentifier’ target=’DIN-561.12’>p.21</ref>	A.10.20	◆	Reference Source(s) of the term.
<termCompList type=termElement>			
<transacGrp>		◆	
<transac type=transactionType>origination</transac>		◆	
<transacNote type=’responsibility’>R. Smith</transacNote>	A.10.02.02.01	◆	Originator an identifier of the person who prepared the word level
<date></date>	A.10.02.01.01	◆	Origination date The date the word level was first created
</transacGrp>		◆	
<transacGrp>		◆	
<transac type=transactionType>creation</transac>		◆	
<transacNote type=’responsibility’>Smith</transacNote>	A.10.02.02.02	◆	Inputter An identifier of the person who types in the information
<trnsacGrp>		◆	

Tag & sample	ISO 12620	Required	Description
<transacGrp>			
<transac type=transactionType>modification</transac>			
<transacNote type='responsibility'>Clarck</transacNote>	A.10.02.02.03		Updater the person having made the latest changes to the information at word level
<date></date>	A.10.02.01.03		Modification date The date when the latest changes to the word level were made
</transacGrp>			
<termCompGrp>			
<termComp>transition</termComp>	A.02.08.02		Term element a particular word that forms part of a term
<termNote type=partOfSpeech>noun</termNote>	A.02.02.01		Part of speech
<termNote type=grammaticalNumber>singular</termNote>	A.02.02.03		Grammatical number
<termNote type=grammaticalGender>masculine</termNote>	A.02.02.02		Grammatical gender
<termCompList type=morphologicalElement>some other morph info</termCompList>	A.02.08.01		Morphological element
<termNote type=pronunciation>…</termNote>	A.02.05		Pronunciation Pronunciation information like accentuation of syllables
</termCompGrp>			
...	…		Other word level items follow here
</termCompList>			
</termGrp>			
</ntig >			
...	…		Other terms follow here
</langSet>			
...	…		Other language level records follow here

Tag & sample	ISO 12620	Required	Description
</termEntry>			
</body>			
<back>			
<refObjectList type=bibl>			Description of the references used in file
<refObject id=piggott97>			Reference object with its identifier
<itemSet type=article>			Type of the reference
<item type=title>Glossary</item>			Title of the reference
</itemSet>			
<itemSet type=author>			
<item type=surname>Piggott</item>			Last name of the author
<item type=fname>Hugh</item>			First name of the author
</itemSet>			
<itemSet type=book>			Type of the reference source
<item type=title>Windpower workshop</item>			Title of the source
<item type=edition>First</item>			Edition of the source
<item type=isbn>1 898049 13 0</item>			ISBN of the source
</itemSet>			
<item type=pages>138-144</item>			Pages of the source
<item type=date>1997-05</item>			Date of the source
<itemSet type=pubname>			Publisher information
<item type=orgName>The Centre for Alternative Technology</item>			Publisher organization name
</itemSet>			
</refObject>			
...			Other reference objects follow here
</refObjectList>			
</back>			

Tag & sample	ISO 12620	Required	Description
`  </text>`			
`</martif>`			

Appendix C. References from EuroTermBank project documentation

This section contains a list of the most important reference materials used while researching and developing the EuroTermBank project.

Literature

Arppe, A., (1995): *Term Extraction from Unrestricted Text*, NoDaLiDa-95.

Auksoriūtė, A. *Terminologijos darbas Lietuvių kalbos institute / Terminology Work at the Institute of the Lithuanian Language // Pabaltijo tautų terminologijos problemos ir Europos Sąjunga / Problems of Baltic States Terminology and the European Union*. Vilnius, 2005. P. 11-20.

Auksoriūtė, A., Gaivenytė, J., Umbrasas, A. *The State of Lithuanian Terminology* // Terminoloģijas Jaunumi: Starptautiskais terminoloģijas seminārs „Terminoloģijas darbs Latvijā, Lietuvā un Igaunijā: sadarbības ieceres ES dalībvalstu kontekstā". Riga. 2003. P.16-25.

Ball, S., D. Rummel (2001) *The IATE Project – towards a Single Terminology Database*: In Proceedings of the 23rd International Conference on Translating and the computer, London, November 2001.

Baums, A., Borzovs, J., Gobzemis, A., Fricnovičs, G., Ilziņa, I., Skujiņa V. *Angļu-latviešu-krievu informātikas vārdnīca – datori, datu apstrāde un pārraide*, 2001, 660 lpp. (English-Latvian-Russian Dictionary of Informatics – Data Processing and Transmission, Riga, 2001, 660 p.).

B*erne Convention for the Protection of Literary and Artistic Works* of September 9, 1886, completed at Paris on May 4, 1896, revised at Berlin on November 13, 1908, completed at Berne on March 20, 1914, revised at Rome on June 2, 1928, at Brussels on June 26, 1948, at Stockholm on July 14, 1967, and at Paris on July 24, 1971, and amended on September 28, 1979 (Berne Convention).

Betz A.; Schmitz K.-D. (1999). *The Terminology Documentation Interchange Format TeDIF*. In: Terminology and Knowledge Engineering TKE'99, Innsbruck, Wien, pp. 782-792.

Borzovs, J. *Mazas tautas lielais papilddarbs.* Referāta tēzes II Pasaules latviešu zinātnieku kongresam, Rīga, 2001, 569.lpp. (*Extra-work of a small nation.* Abstract of a paper of II World-wide Congress of Latvian Scientists, Riga, 2001, p.569).

Borzovs, J., Ilziņa, I., Skujiņa, V., Vancāne, I. *Sistēmiska latviešu datorterminoloģijas izstrāde.* LZA Vēstis, Rīga, 2001, 55.sēj. 1./2. num., lpp.83-91. (*Systemic Development of Latvian Computer Terminology*, Proceedings of Latvian Academy of Sciences, Riga, 2001, vol. 55, N 1/2 pp. 79-83).

Budin, Gerhard: *Recht auf Terminologie – Rechte an Terminologien*. In: Arntz, Reiner; Mayer, Felix; Reisen, Ursula. Geistiges Eigentum an Terminologien. Proceedings of the DTT Symposion held 11.-12. September 1992, p. 29 – 41. Cologne: Deutscher Terminologie-Tag e. V. 1993 (Budin 1993).

Calzolari, N., K. Choukri, M. Gavrilidou, B. Maegaard, P. Baroni, H. FersŅe, A. Lenci, V. Mapelli, M. Monachini, S. Piperidis (2004): *ENABLER Thematic Network of National Projects: Technical, Strategic and Political Issues of LRs*. In Proceedings of the Fourth International Conference on Language Resources and Evaluation, p.937-940, Lisboa.

Castellví, M. T. C., Bagot, R. E., Palatresi, J. 2001. *Automatic term detection: A review of current systems.* In: Bourigault, D., Jacquemin, C. and L'Homme, M.-C. (eds.): *Recent Advances in Computational Terminology.* Amsterdam–Philadelphia: John Benjamins. 53–88.

Drezen, E. *Internationalization of Scientific-technical Terminology.* (Translated from Russian.) Rīga, LU LVI, 2002. – 72 p.

Dudlauskienė, N. *Lietuviškų technikos terminų standartizacija / Standardization of Lithuanian Technical Terms* // Pabaltijo tautų terminologijos problemos ir Europos Sąjunga / Problems of Baltic States Terminology and the European Union. Vilnius, 2005. P. 21-33.

Lenoch Universal Classification System, LUC – vol. 2, Main part. 2. edition. Upd. H. Wellenstein.

Gaivenis, K. *Bendrieji lietuvių terminijos kūrimo ir norminimo principai* // Lietuvos TSR MA darbai. Serija A. Vilnius, 1979. V. 3(68). P. 77-86.

Gaivenis, K. *Kalbiniai ir loginiai terminų reikalavimai / Linguistic and logical requirements of terms //* Terminologija. Vilnius, 1996. V. 3. P. 26-38.

Galinski Ch. *Semantic Interoperability and Language Resources // Terminology and Content Development.* – Copenhagen: 2005. – p. 11-26.

Galinski, Chr. The economics of terminology in the age of the multilingual information society. // *Terminology and Technology Transfer in the Multilingual Information Society.* Proceedings of the 2nd International Conference on Terminology in commemoration of E. Drezen's 110th Anniversary. Vienna/Riga, IITF-Infoterm/LLI of LU, 2003. – P.33-42.

Galinski, Christian; Goebel, Jürgen W. *Guide to Terminology Agreements.* Vienna: TermNet, Internat. Network for Terminology 1996 (Guide 1996).

Galinski, Christian; Wright, Sue Ellen. *Terminology and Copyright.* In: Budin, Gerhard; Wright, Sue Ellen (eds.). Handbook of Terminology Management. Volume I : Basic Aspects of Terminology Management. Amsterdam/Philadelphia: John Benjamins 1997 (Galinski 1997).

Gavrilidou, M., V. Giouli, E. Desipri, P. Labropoulo (ILSP), *Intera – Report on the Multilingual Resources Production*, Deliverable 5.2, e-content EDC-22076 INTERA / 27924.

Goebel, Jürgen W. *Terminologienschutz nach Urheber- und Wettbewerbsrecht.* In: Arntz, Reiner; Mayer, Felix; Reisen, Ursula. Geistiges Eigentum an Terminologien. Proceedings of the DTT Symposion held 11. – 12. September 1992, p. 29 – 41. Cologne: Deutscher Terminologie-Tag e. V. 1993 (Goebel 1993).

Henriksen L., Povlsen C., Vasiljevs A. 2005. *EuroTermBank – a Terminology Resource based on Best Practice.* In Proceedings of LREC 2006, the 5th International Conference on Language Resources and Evaluation, Genoa, on CD-ROM.

Ilziņa, I. *Latviešu termini Eiropas Savienības datortīklos.* Referāta tēzes starptautiskai konferencei *Latviešu grāmata un bibliotēka: 1525.g – 2000.g.*, Rīga, 2000, lpp. 202 – 205. (*Latvian terms in European Union networks.* Abstract of a paper of Intern. Conf. *Latvian Book and Library from 1525 – 2000*, Riga, 2000, pp.202 – 205).

Interinstitutional Committee for Translation and Interpretation, IATE Data Management Group (2003): *Writing rules for the IATE database*, internal report.

Ivanauskienė, *A. Lithuanian Bank of terms – a New Tool for the Regulation of Terminology* // Programme of the International seminar Problems and Tasks of Estonian, Latvian and Lithuanian Terminology in the European Union. Vilnius, 2004, October 5-6.

Johnson I., Macphail A. *IATE – Inter-Agency Terminology Exchange: Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union* // Workshop on Terminology Resources and Computation 2000, LREC 2000 – Athens: 2000.

Johnson, I., A. MacPhail (2000). *IATE – Development of a Single Central Terminology Database for the Institutions and Agencies of the European Union.* In LREC Workshop on Terminology Resources and Computation, Athens.

Keinys St., Auksoriūtė A., Labanauskienė S. Lithuanian terminology on the eve of new century // *Terminology and Technology Transfer in the Multilingual Information Society.* Proceedings of the 2nd International Conference on Terminology in commemoration of E. Drezen's 110th Anniversary. Vienna/Riga, IITF-Infoterm/LLI of LU, 2003. P. 59-64.

Keinys, St. *Terminologijos raida, būklė, vaidmuo ir uždaviniai į XXI amžių įžengus / The State, Development, Role and Tasks of Terminology at the Beginning of the 21th Century* // Vilnius, 2004. P. 219-239.

Keinys, St. *The Development, State and Topicality of Present Day Lithuanian Terminology* // Pabaltijo tautų terminologijos problemos ir Europos Sąjunga / Problems of Baltic States Terminology and the European Union. Vilnius, 2005. P. 48-67.

Kis, B., Villada Moirón, B., Bíró, T., Bouma, G., Pohl, G., Ugray, G., Nerbonne, J. 2004b. *Methods for the Extraction of Hungarian Multi-Word Lexemes.* In: Decadt, B. (ed.): *Proceedings of CLIN-2003.* Antwerp: University of Antwerp.

Kis, B., Villada, B., Bouma, G., Bíró, T., Nerbonne, J., Ugray, G. and Pohl, G. 2004a. *A new approach to the corpus-based statistical investigation of Hungarian multi-word lexemes*, In: Proceedings of LREC 2004. Lisbon.

Konferences „Zinātnes valoda"materiāli. (Proceedings of the conference „Language of Science".) Rīga, National Language Commission, 2003 – 51 lpp.

Krūmiņa, V., Skujiņa, V. *Normatīvo aktu izstrādes rokasgrāmata. (Manual for creating normative acts.)* Rīga, Valsts kanceleja, 2002. – 116 lpp.

Latviešu valoda – robežu paplašināšana. (The Latvian Language – Extending the borders.) Rīga, National Language Commission, 2005. – 216 lpp.

Laurent Romary: *An abstract model for the representation of multilingual Terminological data: TMF – Terminological Markup Framework.* Published on: http://www.papillon-dictionary.org/info_media/1620632.pdf.

Maslauskienė, R. *Terminų standartizavimas / Standardization of Terms* // Terminologijos istorijos ir dabarties problemos. Vilnius, 2004 (256-270).

Melby, Alan K.; Schmitz, Klaus-Dirk; Wright, Sue Ellen: *The Machine Readable Terminology Interchange Format (MARTIF) — Putting Complexity in Perspective.* In: TermNet News, Nr. 54/55-1996, S. 11-21.

Meškauskienė, S. *Matematikos ir informatikos instituto lietuvių kalbos terminų bazė (Lietuvių kalbos terminynas) / The Lithuanian Language Terms Base of the Institute of Mathematics and Informatics //* Pabaltijo tautų terminologijos problemos ir Europos Sąjunga / Problems of Baltic States Terminology and the European Union. Vilnius, 2005. P.89-99.

Piccioni, Lorenzo, Eros Zanchetta (2004). *XTERM: A Flexible Standard-Compliant XML-based Termbase Management System.* In Proceedings of LREC, the IV International Conference on Language Resources and Evaluation. Lisboa, pp. 469-473.

Picht, H., Draskau, J. *Terminology: An introduction.* Copenhagen, School of Economics; University of Surrey England, 1985. – 265 p.

Quality and Reliability SA (2000), System analysis and design, report of the IATE project.

Rat für Deutschsprachige Terminologie. Terminologie und Urheberrecht. Ms. Workshop 10. – 11. November 2000 (RaDT 2000).

Rummel, D., S. Ball (2001). *The IATE Project – Towards a Single Terminology Database for the EU.* In Proceedings of ASLIB 2001, the 23rd International Conference on Translation and the Computer, London.

Sager, J.C. (1990), *A Practical Course in Terminology Processing.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Schmitz, Klaus-Dirk: *Die neuen Terminologiedatenbanken: online statt offline.* In: Felix Mayer, Klaus-Dirk Schmitz, Jutta Zeumer (eds.): Terminologie und Wissensmanagement. Akten des Symposions, Köln, 12.-13. April 2002. Köln: Deutscher Terminologie-Tag e.V.

Schmitz, Klaus-Dirk: *MARTIF – Ein SGML-basiertes Austauschformat für terminologische Daten.* In: Wiebke Möhr, Ingrid Schmidt (eds.) (1999): SGML und XML – Anwendungen und Perspektiven. Berlin: Springer, S. 109-121.

Schmitz, Klaus-Dirk: *Systeme zur Terminologieverwaltung.* In: Technische Kommunikation, No. 2/2001, p. 34-39.

Schmitz, Klaus-Dirk; Galinski, Christian (eds.). *GTW-Report. Guidelines for the Design and Implementation of Terminology Data Banks.* Association for Terminology and Knowledge Transfer 1996 (GTW-Report 1996)

Schmitz; Klaus-Dirk: *MARTIF – A New ISO Standard for the Interchange of Terminological Data.* In: TermNet News, Nr. 50/51-/1995, S. 4-5.

Situācijas izpēte latviešu terminoloģijas izstrādes, saskaņošanas un apstiprināšanas jomā: problēmu identifikācija un to risinājumi. (Investigation of the situation in the terminology elaboration, harmonization and approval of the Latvian Terminology: Problem identification and solution.) TTC. Rīga, 2004.

Skadiņš R., Vasiljevs A. 2004. *Multilingual Terminology Portal – termini.letonika.lv.* In: Proceedings of First Baltic Conference „Human Language Technologies – the Baltic Perspective", Riga, 183-186.

Skujiņa, V. *Gramatikas kategoriju loma terminu sistēmas veidošanā. (The role of grammar categories for the developing of term system.)* // Baltu Filoloģija, XI *(2).* Rīga, LU, 2002. – P. 67-78.

Skujiņa, V. *Latīņu un grieķu cilmes vārddaļu vārdnīca. (Dictionary of the Latin and Greek elements.)* Rīga, Kamene, 1999. – 233 lpp.

Skujiņa, V. *Latviešu terminoloģijas izstrādes principi. (The Principles of Formation of Latvian Terminology.)* 2nd ed. Rīga, LZA/LU LVI, 2002. – 224 p.

Skujiņa, V. *Latviešu valodas terminoloģijas attīstība nacionālo un internacionālo tendenču mijiedarbībā. (Development of Latvian terminology in the framework of interactive national and international tendencies.)* // Baltu Filoloģija, *X.* Rīga, LU, 2001. – P. 113-125.

Skujiņa, V. *The Specificity of the Term and the Concept of the Termeme* // IITF Infoterm Multiligualism in Specialist Communication. Vienna, TermNet, 1996. – Vol.2. – P. 1123-1130.

Stellbrink, Hans-Jürgen. *Urheberrechte an Terminologien – ein klares "Jein".* In: Arntz, Reiner; Mayer, Felix; Reisen, Ursula. Geistiges Eigentum an Terminologien. Proceedings of the DTT Symposion held 11. – 12. September 1992, p. 1 – 11. Cologne: Deutscher Terminologie-Tag e. V. 1993 (Stellbrink 1993).

Streiter, O., D. Zielinski, I. Ties, L. Voltmer (2003): *Term Extraction for Ladin: An Example-based Approach*, TALN 2003.

Suonuuti, H. (1997), *Guide to Terminology.* Helsinki:Tekniikan Sanastokeskus.

Terminology and Technology Transfer in the Multilingual Information Society. Proceedings of the 2nd International Conference on Terminology in commemoration of E. Drezen's 110th Anniversary. Vienna/Riga, IITF-Infoterm/LLI of LU, 2003. – 184 p.

Thurmair, G. (2003): *Making Term Extraction Usable*, EAMT workshop 2003, Dublin (Powerpoint presentation).

Translation Handbook for Latvian Legislation. (lv-en) Rīga, TTC, 2000. – 80 p.

Vancāne, I., Borzovs, J. *Development of the Latvian Terminology of Information Technologies.* Abstract of a paper of Intern. Conf. & Exib. *Information Technologies and Telecommunications in the Baltic States*, Riga, 1999, pp. 98-102.

Vasiļjevs A., Skadiņš R. (2005). *Eurotermbank terminology database and cooperation network.* In Proceedings of the Second Baltic Conference on Human Language Technologies, Tallinn, pp. 347-352.

Veddern, Michael. *Die Durchsetzungs-Richtlinie 2004/48/EG – Ein weiterer Schritt zur Harmonisierung der geistigen Eigentumsrechte in Europa.* In: IPR Helpdesk Bulletin Nr.16, Aug. – Sept. 2004 (Veddern 2004).

Weissenhofer, P. *Conceptology in terminology theory, semantics and word-formation.* Wien, TermNet, 1995. – 271 p.

Wright, Sue Ellen (2005). *A Guide to Terminological Data Categories – Extracting the Essentials from the Maze.* In Proceedings of TKE 2005, the 7th International Conference on Terminology and Knowledge Engineering. Copenhagen, pp. 63-77.

Wright, Sue Ellen. *Intellectual property rights and terminology management.* In: TermNet News 52/53 1996 (Wright 1996).

Standards and EU directives

ISO 704:2000 Terminology work – Principles and methods

ISO 860:1996(E) Terminology work – Harmonization of concepts and terms

ISO 10241:1992(E) International terminology standards – Preparation and layout

ISO 12200:1999 – Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) — Negotiated interchange

ISO 12620:1999 – Computer applications in terminology – Data categories

ISO 16642:2003 – Computer applications in terminology – Terminological markup framework (TMF)

Directive 91/250/EC of the European Parliament and of the Council of 14 May 1991 on the legal protection of computer programs (Computer Program Directive)

Directive 92/100/EC of the European Parliament and of the Council of 19 November 1992 on rental right and lending right and on certain rights related to copyright in the field of intellectual property (Rental Right Directive)

Directive 93/83/EC of the European Parliament and of the Council of 27 September 1993 on the coordination of certain rules concerning copyright and rights related to copyright applicable to satellite broadcasting and cable retransmission (Broadcasting Directive)

Directive 93/98/EC of the European Parliament and of the Council of 29 October 1993 harmonizing the term of protection of copyright and certain related rights (Copyright Directive 1)

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive)

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (E-commerce Directive)

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society (Copyright Directive 2)

Directive 2004/48 EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights (IPR Enforcement Directive)

Internet resources

The Internet resources provided in this section are available and relevant at the time of preparing this publication.

Council for the Polish Language (Rada Języka Polskiego):
http://www.rjp.pl

Deutsches Terminologie-Portal:
http://www.iim.fh-koeln.de/dtp

DIN, Germany:
http://www.beuth.de
http://www.din.de

Enabler:
http://www.elda.org/article103.htm
http://www.enabler-network.org

Estonian Legal Language Centre:
http://www.legaltext.ee

Estonian Terminology Association:
http://www.eter.ee

Eurovoc:
http://eurovoc.europa.eu

Infoterm:
http://www.infoterm.info

Institute of the Estonian Language:
http://www.eki.ee

Institute of the Lithuanian Language:
http://www.lki.lt

Intera:
http://www.elda.org/rubrique22.html

International Organization for Standardization:
http://www.iso.org

ISO policies and procedures for copyright exploitation rights and sales of ISO publications:
http://isotc.iso.org/livelink/livelink/fetch/2000/2489/186491/186621/802824/POCOSA.pdf

IPR Helpdesk:
http://www.ipr-helpdesk.org

Latvian Terminology portal:
http://www.termnet.lv

Lenoch:
http://www.uibk.ac.at/translation/termlogy/lenoch.html
http://www.disclic.unige.it/certem/arc/lenoch.pdf

Lithuanian Language Term Base:
http://www.terminynas.lt

Lithuanian Standards Board:
http://www.lsd.lt

Polish Society of Economic, Legal and Court Translators TEPIS (Polskie Towarzystwo Tłumaczy Ekonomicznych, Prawniczych i Sądowych TEPIS):
http://www.tepis.org.pl

Polish Standardization Committee 'PKN' (Polski Komitet Normalizacyjny):
http://www.pkn.pl

Polska Terminologia Informatyczna – Biuro Tłumaczeń Informatycznych:
http:// www.btinfo.pl

Standards-based Access to multilingual Lexicons and Terminologies (SALT):
http://www.iim.fh-koeln.de/iim/salt.html
http://www.loria.fr/projets/SALT/saltsite.html
http://www.ttt.org/salt/description.html

State Commission of the Lithuanian Language:
http://www.vlkk.lt

TBX Standard:
http:// www.lisa.org/standards/tbx/

The Office of the European Integration Committee 'UKIE' (Urząd Komitetu Integracji Europejskiej):
http://www.ukie.gov.pl

The PAVEL Terminology Tutorial:
http://www.termium.gc.ca/didacticiel_tutorial/english/lesson1/index_e.html

World Intellectual Property Organization:
http://www.wipo.int

Consistent, harmonized and easily accessible terminology is an extremely important stronghold for ensuring true multilingualism in the European Union and throughout the world. From legislation and trade to the needs and mobility of each individual, terminology is the key for easy, fast and reliable communications. The rapid path of changes in technologies and the global economy leads to ever growing introduction of new concepts and terms to describe them. Globalization from the one side and growing language awareness from the other side dictate the need to consolidate national terminology resources, harmonize international terminology, and provide online access to reliable multilingual terminology.

This publication summarizes the experiences and findings of the EuroTermBank project, part of the European Union eContent program. It is aimed at individuals and organizations interested and involved in all aspects of terminology management.

eContent

EuroTermBank Consortium
www.eurotermbank.com

ISBN 9984-9133-4-1