Igor Kudashev

QUALITY ASSURANCE
IN TERMINOLOGY MANAGEMENT

Recommendations from the TermFactory project

Igor Kudashev holds a PhD degree and the title of associate professor from the University of Helsinki, Finland. He is an author or co-author of several published terminological dictionaries, including the Finnish-Russian Forestry Dictionary, which was awarded the International Terminology Award for Applied Research and Development by the European Association for Terminology (EAFT) in 2008. The author has also developed MyTerMS, an in-house terminology management system. Experience obtained in these projects proved very valuable when addressing such a challenging topic as quality assurance in terminology management and collaborative content creation.

This guide provides recommendations on quality assurance in terminology management as well as practical quality assurance tools from the TermFactory research and development project. Launched at the University of Helsinki in 2008, the TermFactory project is developing an ontology-underpinned platform for collaborative terminology work. The guide is addressed to a wide range of experts engaged in terminology work and terminology management.

UNIVERSITY OF HELSINKI

Igor Kudashev

QUALITY ASSURANCE
IN TERMINOLOGY MANAGEMENT

Recommendations from the
TermFactory project

Helsinki 2013

Igor Kudashev

QUALITY ASSURANCE IN TERMINOLOGY MANAGEMENT:
RECOMMENDATIONS FROM THE TERMFACTORY PROJECT

Cover design: Igor Kudashev and Hanna Sario

Back cover photograph by Irina Kudasheva

TABLE OF CONTENTS

Preface

This guide provides recommendations on quality assurance in terminology management as well as practical quality assurance tools from the *TermFactory* research and development project. *TermFactory* is the name of a platform developed for presentation of collaboratively created terminological, ontological and lexicographical content in the form of ontologies. The platform has been developed since 2007 by professor Lauri Carlson (University of Helsinki, Department of Modern Languages) and his team within a number of projects, the most important of which are listed below:

- The *ContentFactory* project (2008–2011) financed by the Finnish Funding Agency for Technology and Innovation (Tekes), the city of Kouvola and a number of language industry companies. Among other things, the project set out to develop a working demo of the *TermFactory* platform.

- *ContentFactory*'s follow-up project *CF-PreComm* (2011–2013) aimed at preparation of the research results obtained in the *ContentFactory* project for commercialization.

- The *Mobster* project (2011–2013) financed by Tekes and a number of health care companies. The goal of the project was to develop and test a mobile and integrated dictation and communication application for the health care sector. The *TermFactory* platform was a part of the terminology management solution for the project.

- The *MOLTO* (Multilingual Online Translation) project (2010–2013) financed by the EU Seventh Framework Programme. The goal of the project was to develop a machine translation system capable of providing publishing-quality automatic translations in selected domains. The *TermFactory* platform was the primary tool for storing and managing linguistic, terminological and ontological data required by the translation system.

In addition to developing ontology-based technology for managing lexical data, the *TermFactory* project also creates a methodology and quality assurance

instruments for collaborative terminology work. This part of the project has primarily been the responsibility of researchers from the Palmenia unit of the University of Helsinki (Igor Kudashev and Irina Kudasheva), who have theoretical and practical expertise in both terminology work and terminology management.

This guide is not intended to provide a comprehensive theoretical framework for quality assurance in terminology work and terminology management. Development of such a framework is a topic for a whole new project. Instead, we provide the results of the bottom-up research deriving from the needs of the specified project and application. However, we also provide a theoretical basis for our recommendations in most cases and hope that they will have both practical and theoretical value for various experts engaged in terminology work and terminology management.

Not all of the recommendations in this guide have actually been implemented in the *TermFactory* platform. The guide reflects the author's point of view, and the recommendations we provide are partly *from* the *TermFactory* project and partly *for* it. Readers should also keep in mind that the guide reflects the state of the art at the time of publishing, while the development of the *TermFactory* platform and related methodology continues.

The author will be grateful for feedback from readers on this guide, as it will help him prepare a more comprehensive publication on the subject in the future. Please send your comments to Igor Kudashev (igor.kudashev [at] helsinki.fi).

Acknowledgements

1   Overview of the *TermFactory* platform

In order for the readers to better understand the goals and requirements of the *TermFactory* project as regards quality assurance, we start with a brief overview of the *TermFactory* platform. A much more detailed description of the system's architecture and the philosophy behind it is available at http://www.helsinki.fi/~lcarlson/CF/TF/doc/TFManual_en.htm (best viewed with the *FireFox* browser).

1.1 *Ontology-based architecture*

The *TermFactory* platform differs from other terminology management systems in several respects. The first major difference, which accounts for several other peculiarities, is that *TermFactory* is an ontology-based system.

Ontologies are formal representations of entities and relations between them. Unlike hierarchical databases and plain XML that use the 'container inside container' approach, ontologies are semantic networks. In ontologies, all nodes are equal, and any node can be taken as root or focus, which allows the user to freely choose the topic and aspect of description. Each node has a unique identifier (URI), which allows one to address nodes directly and not via their parent elements. Since identifiers are also global, the node can be unambiguously referred to in external resources.

Ontologies have become a significant format for describing complex concept systems in many areas, especially in natural sciences and medicine. The main focus in ontology work has been on concepts. Natural language terminology is usually included only as simple string labels and not as ontological resources in their own right. However, nothing prevents us from treating terms and their descriptions as ontology resources as well.

TermFactory comprises an ontology schema, a web API, and a platform for collaborative terminology work that is based on the explicit ontological representation

of concepts and designations as well as their classes and properties as points in a semantic network.

Terminology and lexicography traditionally differ in the choice of the main object of description. In terminology, it is usually the concept (see Terminological Markup Framework – ISO 16642:2003) while in lexicography it is usually the lemmatized expression (see Lexical Mark-up Framework – ISO 24613:2008). The *TermFactory* platform is neutral between the two orientations, describing meanings and forms as independent ontology resources. A given meaning (concept) and form (designation) are linked by a word sense (term) which is also a separate ontology resource.

Among other things, TermFactory representation allows separation of a) the semantic properties of concepts, b) properties that expressions have in virtue of general language grammar, and c) properties that are associated with a term in a domain-specific meaning. Each part of the description can be developed independently and combined at will.

For example, ontology engineers may develop a monolingual or language-neutral ontology, terminologists may supplement it with multilingual terminology and semantic description in a natural language, and linguists may add the linguistic description. However, the workflow may start from any of these levels.

Because of its flexibility as regards the object of description, *TermFactory* can easily accommodate different kinds of lexical resources – not only concept-oriented and lemma-oriented but also more specific ones, such as synset-oriented *WordNet*. In particular, the Finnish translation of *WordNet* 3.0. obtained from the FIN-CLARIN project was successfully converted into the TermFactory format. This means that in spite of its original name and intention, the *TermFactory* platform is growing into a more generic system capable of managing various types of lexical and ontological resources.

Ontologies require that the input data be expressed precisely and explicitly but they also guarantee the same degree of precision and explicitness when the data is passed on to other agents. Data presented in the form of an ontology can be directly

used by NLP applications, such as machine translation systems, speech recognition and synthesis applications, computer assistants and the like. Ontologies can be automatically checked for logical errors, which substantially improves the overall quality of the data. Ontologies are also equipped with tools for automatic reasoning about relations which have not been explicitly stated but which are logically deducible. This reduces the volume of manual work, especially when preparing lexical content for use in NLP applications.

Being an undirected graph, TermFactory term ontology is not bound to a fixed notion of an entry as the 'container' that keeps individual data elements together. An 'entry' in the TermFactory approach can be conceived of as the result set of an ontology query. Static entries are replaced in the TermFactory model by dynamic views that result from different traversals and serializations of the term ontology graph. The same data can be parcelled up and presented to the end users in many different ways. *TermFactory* thus contributes to the implementation of the idea of dynamic and customizable electronic dictionaries that have been on lexicographers' desiderata list for a long time (e.g., Atkins 2002; Varantola 2002; de Schryver 2003).

1.2 Support for collaborative content creation

TermFactory has been conceived of as a platform for collaborative terminology work from the very start. At the outset of the project, the plan was to develop specific dedicated tools for querying, browsing, editing and discussing terms. However, the focus has since shifted to providing plug-ins and mash-ups that can be embedded in a variety of present or future third-party platforms. In particular, *MediaWiki*, *Drupal* and *Disqus* have been used as testbeds.

Currently, the *TermFactory* platform does not support multi-user editing of term ontologies with real time shared access, record level locking, etc., as collaborative terminology work is being intended to be carried out in forum/wiki platforms with TermFactory plug-ins. Once verified by the community and professional

terminologists, the data is moved to the read-only section of the repository, where it gets updated through batch updates.

1.3 A distributed and cloud-ready system

As a web-based service, *TermFactory* is a distributed resource 'in the cloud'. In practice this means that data is not stored in one place but is distributed between multiple repositories of various types.

TermFactory consists of a third-party collaborative forum and wiki front-end servers and TermFactory repository back-end servers. The front-ends communicate with the back-ends through TermFactory mash-ups for querying, editing, etc. The back-end servers communicate with each other through web service requests and responses. The repositories and the collaboration platforms may or may not share infrastructure at some level. The two subsystems can also work independently of each other.

TermFactory's user management is currently based on the user management of the *GlobalSight* open-source translation management system. *TermFactory* users are managed through a *GlobalSight* server instance with TermFactory-specific extensions. *GlobalSight* user management is based on a registry of users, companies, locale pairs, activity types and roles. *TermFactory* adds one more attribute to the list, called a domain. A TermFactory domain is a regular expression on ontology URIs that specifies which parts of the TermFactory term ontology are available to each company, role and user.

2   Quality assurance infrastructure

2.1 Definition of basic concepts

In this section, we tailor the definitions of some basic concepts (such as *terminology management, term bank, LSP designation, quality*, *quality assurance, metadata*) for the purposes of this guide, thus restricting its scope.

Terminology management in the broad sense is 'any deliberate manipulation of terminological information' (Handbook of Terminology Management 1997: 1). For the purposes of this guide, we narrow this concept to cover the management of *terminological descriptions* in *term banks*.

By *terminological descriptions*, we mean information about the form, meaning, relations and usage of *LSP designations*. In terminological reference resources like term banks, terminological descriptions provide information that is supposed to help users understand, use or replace LSP designations in external contexts (cf. Kudashev 2007: 47–48).

LSP designation is an umbrella concept for various types of lexicalized object in languages for special purposes. In addition to terms, LSP designations also include proper names, nomenclature, set phrases, etc. Please refer to *Section 3.2* for an extended discussion of what may constitute an object of description in a term bank.

As was mentioned above, since in the *TermFactory* platform, forms and meanings can be objects of description in their own right, it is fair to speak about *description of the form* and *semantic description* as separate objects and not just parts of *terminological description*. However, our primary focus is on terminological descriptions as a whole.

A *term bank* can be defined as a totality of structured terminological data and a framework for its management (cf. definitions in ISO 1087-2:2000: 12). This framework consists of metadata and software required for managing terminological data.

Metadata ('data about data') is a broad and complex concept, the intention of which varies in different communities. For our purposes, the following definition of metadata may be adopted: 'structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource' (NISO 2004: 1).

For *quality* and *quality assurance*, we adopt definitions from ISO 8402:1994, an international standard on quality management and quality assurance. This standard defines *quality* as the 'totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs' (ISO 8402:1994: 23), and *quality assurance* as 'all the planned and systematic activities implemented… to provide adequate confidence that an entity will fulfil requirements for quality' (ISO 8402:1994: 15–16). The ISO standard also stresses the relative and dynamic character of quality and the fact that the degree of quality is ultimately determined by the users and applications of the product or service in question.

Quality assurance is implemented in term banks with the help of *quality assurance infrastructure* – physical and organizational structures and facilities designed to assure the specified quality level of terminological data and terminology management operations. Elements of quality assurance infrastructure are the main topic of this guide.

2.2 Classification of elements of quality assurance infrastructure

Quality assurance infrastructure in a term bank mostly consists of various types of metadata. Depending on its function and the object to which it relates, metadata can be divided into *structural* and *descriptive metadata*. Another important element of quality assurance infrastructure is *methodological data*. In the following sections, we discuss this classification in more detail.

2.2.1 Methodological data

Methodological data is the 'how-to' data which is not directly related to terminological descriptions or data containers. It includes guidelines for content creators, users, administrators and developers of a term bank. The guidelines may relate to content creation, data organization or data management, and may include recommendations on the general methodology of terminology work and terminology management, issues specific to a particular method of terminology work, as well as interaction with the terminology management system.

2.2.2 Structural metadata

Structural metadata comprises semantic, syntactical and value domain specifications of data classes. A typical example of structural metadata is a *data category schema*, according to which terminological descriptions in a term bank are split into blocks. Depending on the type, organization and complexity of the schema it may also be referred to as a data category inventory, data element set, registry, repository, catalogue, dictionary, ontology, etc. In particular, a data category schema may contain the following types of data:

- data category set, i.e., a set of classes into which terminological descriptions are split
- relations between data categories (e.g., the data category 'gender' is a hyponym of the hyperonym 'grammatical information' which in its turn is a subclass of the class 'information about the form')
- syntactical restrictions (e.g., the data category 'gender' cannot appear together with the data category 'declination')
- value type and value domain restrictions (e.g., the data category 'update date' can only accept dates in a particular date format, or the data category 'declination' can only accept the values 'I' and 'II').

Value domain restrictions may be specified by a pattern or enumeration, e.g., the data category 'language code' may be specified as only accepting values listed in the ISO 639-1:2002 standard ('Codes for the Representation of Names of Languages').

In this way, structural metadata may be compared to a template according to which terminological descriptions are structured, organized and validated. It does not relate to terminological descriptions as such but to *classes of data* – data categories.

Data categories are manifested in a term bank in containers usually called *data fields* (ISO 1087-2:2000: 13). Data fields may form larger persistent or virtual entities (such as sections, entries, collections and their subsets/supersets), the character and nomenclature of which depend on the type of terminological resource and the technical implementation of the underlying terminology management system. We call these data containers *terminology management entities*.

2.2.3   Descriptive metadata

Descriptive metadata relates to terminological descriptions and/or terminology management entities. Drawing the line between the two types of descriptive data is sometimes problematic in practice.

Descriptive metadata helps users interpret, evaluate and improve terminological descriptions as well as navigate within terminology management entities and administer them. Below are a few examples of descriptive metadata:

- explanations on terminographic conventions used in a particular terminological collection
- principles of compilation of a terminological collection
- information on the written and oral sources of a particular collection
- working comments and users' notes on terminological descriptions
- rating of terminological descriptions or their parts
- administrative data, such as authorization information

- data field labels and other means that help users locate and identify terminology management entities
- disambiguation means that help in choosing between formally identical topics of description (such as homonymy indexes or disambiguation notes)
- entity-level layout (e.g., style settings applicable to all data containers of a particular type)
- local layout (e.g., local exceptions to the generic style rules, or inline formatting).

Sometimes it is not easy to distinguish between terminological descriptions and descriptive metadata. As 'no data is always metadata' (ISO/IEC 11179-1:2004: 10), it is impossible to identify metadata just by looking at it (Bargmeyer and Gillman 2000: 1). Data becomes metadata under particular circumstances, for particular purposes, and with certain perspectives, called the *context* (ISO/IEC 11179-1:2004: 10).

For example, language identifiers (*en*, *de*, etc.) may be considered an indispensable part of terminological descriptions as they indicate to which language the described LSP designation belongs. At the same time, if they are also used by the system for the processing of data, e.g., for selecting the correct spellchecker, then they also belong to administrative data.

Data category *source* may also have various functions. For example, if bibliographical reference is supplied in order to describe the term as a linguistic object (e.g., the source of the first occurrence of a term, or a source of authority for a standardized term), it is a part of terminological description. On the other hand, if the only purpose of a source field is to create a hyperlink to an external corpus, it counts as administrative data. The source field may also combine these two functions or have additional ones.

2.3 Elements of quality assurance infrastructure included in this guide

It is hardly possible to cover all elements of quality assurance infrastructure in a single guide, as requirements for quality assurance are project-specific. In this guide, our focus is on those aspects of quality assurance that we considered particularly important for the implementation of the *TermFactory* platform within the scope of the *ContentFactory* project.

As methodology has an impact on other parts of quality assurance infrastructure in a term bank, we start by discussing several methodological issues. First, we consider workflows and user roles in collaborative terminology work as opposed to traditional models of content creation. We discuss the benefits and challenges of the collaborative approach and make suggestions for maximizing the benefits and minimizing the risks. We also present our vision of the general principles of collaborative terminology work and a possible 'ecosystem' for it.

Another important methodological question is what may be an object of description in a term bank. We advocate inclusion of a large range of LSP designations into term banks, not only prototypical terms but also appellations, nomenclature, terminoids, term elements and lexicalized LSP units. At the same time, we argue that such units as equations, logical expressions, collocations and standard texts that are classified under the category 'term type' in ISO 12620 (ISO 12620:1999: 8–10) can hardly be objects of terminological description. We also discuss various forms in which objects of description may be stored in a term bank, such as lemmatized, normalized and optimized forms. Finally, we introduce the concept of *terminological lexeme* that allows us to take into account the whole range of LSP designations relevant for terminology management as well as their various forms.

A large part of the manual is devoted to structural aspects of the quality assurance infrastructure. In particular, we consider the two most important characteristics of any LSP designation – language and subject field. The question of

language identification and indication has been basically solved in international standards and guidelines, but there are some nuances that require special attention when designing a term bank. Besides, in some questions, such as customization of collation rules, existing standards are insufficient for lexicographic purposes and need to be extended.

Domain classification is of the utmost importance for term banks as they typically contain terminology from multiple subject fields. Unfortunately, most existing domain classifications, thesauri and ontologies are not immediately suitable for use in terminology management. We discuss the principles of compilation of a core domain classification specifically designed for the purposes of multilingual collaborative terminology work and provide such a classification in four languages in *Appendix 3*.

Data category classification is yet another important element of quality assurance infrastructure in a term bank. The use of different data category schemata in different collections diminishes interoperability and complicates data exchange. As a remedy, we propose a bridging classification of data categories based on the linguistic functions of the data. This classification addresses such terminology management problems as advanced full-entry search organization and customization of entry views in collections with different structures. In *Appendix 1*, we provide the bridging classification of terminological data and a mapping of ISO 12620:1999 data categories onto it.

In the last chapter, we focus on two types of descriptive metadata: indication of sources and administrative data. Indication of sources is an important part of quality assurance in a term bank, as it allows users to estimate the quality of terminological descriptions. In collaborative content creation, the importance of source indication and documentation can scarcely be overestimated. We discuss both minimal requirements and advanced support for the documentation of sources in a term bank and provide templates for documentation of various types of sources in *Appendix 4*.

Since contributors may also be viewed as sources in a collaborative platform, we provide a template for the documentation of contributors' expertise in *Appendix 5*. The template helps evaluate the proficiency of contributors and the reliability of the

data produced by them. Besides, it can be used for seeking people with a suitable background for new projects, administrative tasks, and so on. In the last section, we provide a definition and a multi-facet classification of administrative data, which may help plan data management in terminological resources.

3 Methodological aspects of quality assurance in collaborative terminology work

In this part of the guide, we deal with several methodological questions related to terminology work in general and collaborative terminology work in particular. We discuss the differences between the traditional and collaborative models of terminology work, define the concept of collaborative terminology work and provide recommendations on the principles of collaborative work for the *TermFactory* platform. The second part of the chapter is devoted to the question of what may constitute an object of description in a term bank.

3.1 Workflows and user roles in collaborative terminology work

We start with a brief overview of the methodology adopted in traditional terminology work and then compare it with collaborative content creation. Such a comparison helps us identify the essential characteristics of collaborative terminology work and thus define it. The comparison also allows us to identify the strengths and weaknesses of the collaborative approach and work out a quality assurance strategy for it.

3.1.1 Workflows in traditional terminology work

The general principles of terminology work are described in ISO 704:2009 ('Terminology work – Principles and methods') and in various manuals of terminology work (e.g., Felber 1984; Picht & Draskau 1985; Sanastotyön käsikirja 1989; Sager 1990; Grinev 1993; Handbook of Terminology Management 1997/2001; Cabré 1999; Handbook of Terminology 2001; Suonuuti 2001; Leičik 2006). Most of these guides provide one or more workflow scenarios for terminology work. However, creation of a

universal model for terminology work sounds a difficult if not impossible task because of the variety of goals and requirements in terminological projects as well as the varying backgrounds and preferences of content creators.

While some stages of the workflow logically precede others (e.g., proofreading usually makes sense only when the material is almost ready), the sequence of other stages is largely a matter of preference. Many stages of terminology work are also continuous, iterative or overlapping (e.g., Suonuuti 1997: 16; Nykänen 1999: 63, 68). For example, consultations with domain experts are usually continuous while concept analysis and searching for foreign equivalents often overlap and are circular.

Some stages of terminology work are also optional. For example, sending the glossary to external experts for evaluation is a valuable means of quality assurance but this stage cannot be considered compulsory for every terminological project.

Traditionally, terminology work has been done in a centralized way by relatively small teams of language professionals and domain experts. Several examples of workflow scenarios used in centralized terminology work are provided below. The first is a general model (Leičik 1975: 12–13). The second one represents a workflow used in the preparation of international terminology standards (ISO 10241:1992). The third scenario describes normative terminology work conducted at the national level (Suonuuti 1997: 15–16; Nykänen 1999: 62–63), while the fourth model is an example of a workflow suitable for descriptive bilingual terminology work (Kudasheva 2009: 112–124).

General model of terminology work

- Studying the literature on the subject.
- Term extraction and documentation of context definitions and contexts.
- Checking of term list by domain experts.
- Creation of concept charts by terminologists in cooperation with domain experts.

- Population of concept charts with terms by terminologists in cooperation with domain experts.
- For a translators' glossary: searching for equivalents in the target language.
- For a normative glossary: domain experts defining the concepts in cooperation with terminologists.
- For a normative glossary: evaluation of terms from the point of view of their normative status by domain experts in cooperation with terminologists.
- Revision of the glossary by a board of experts.
- Preparation of the glossary for publication.

Preparation of international terminology standards

- Preliminary work
  - Needs analysis.
  - Defining the target groups.
  - Subject delimitation.
  - Selection and evaluation of sources.
  - Defining the number of concepts.
  - Choosing the languages.
  - Drawing up the schedule.
- Working procedure
  - Collecting terminological data.
  - Recording terminological data.
  - Establishing the term list.
  - Establishing concept fields and concept systems.
  - Formulating definitions.
  - Coining and selecting terms.
- Terminography
  - Defining the entry structure.

- Defining the reference structure.
- Defining the order of entries.
- Choosing the methods of accessing the terminology.
- Adding indexes.
- Choosing the presentation of entries and headwords.

Normative terminology work conducted at the national level

- Defining the target group and the purpose of the glossary.
- Division of the domain into subdomains.
- Defining the scale of the glossary (number of headwords).
- Defining the time scale of the project.
- Critical analysis of existing glossaries in the given domain.
- Selection and evaluation of sources.
- Preliminary term selection.
- Final term selection.
- (Thematic) grouping of terms.
- Narrowing down the scope of the glossary.
- Creation of concept charts.
- Searching for equivalents.
- Making draft definitions.
- Final editing of concept charts.
- Final editing of definitions.
- Checking equivalents.
- Comments from a board of domain experts.
- Final revision of the glossary.
- Preparation of the glossary for publication.

Descriptive bilingual terminology work

- Terminologists: examination of existing dictionaries in the given domain.
- Terminologists and editors: defining the target group and purpose of the dictionary.
- Terminologists and editors: defining the number of concepts.
- Terminologists and editors: defining the time scale of the project.
- Terminologists and editors: writing the dictionary plan.
- Terminologists: division of the domain into subdomains.
- Domain experts: comments on the subdomain classification.
- Terminologists and domain experts: selection of sources.
- Terminologists: term extraction, compilation of draft definitions and notes, concept analysis (including the creation of concept charts) in each subdomain in the source language.
- Source language (SL) domain experts: checking of terms, definitions, notes and concept charts in each subdomain.
- SL editor: checking of terms, definitions, notes and concept charts in each subdomain.
- SL editor and domain experts: negotiations on the changes proposed by the editor.
- SL editor: checking of terms, definitions, notes and concept charts in all subdomains.
- SL editor and domain experts: negotiations on the proposed changes.
- Terminologists: concept analysis, creation of concept charts and searching for equivalents in the target language (TL).
- TL domain experts: checking equivalents, transfer notes and concept charts in each subdomain.
- Chief TL domain expert: checking of equivalents, notes and concept charts in all subdomains.
- TL editor: checking equivalents, notes and concept charts in all subdomains

- TL editor and TL domain experts: negotiations on the changes proposed by the editor.
- SL proof-reader: checking the SL part of the dictionary.
- TL proof-reader: checking the TL part of the dictionary.
- Preparation of the dictionary for publication (Preface, User's Guide, etc.).

3.3.2   Comparison of traditional and collaborative content creation

None of the workflow scenarios described above can be used as such in collaborative terminology work, as the latter is decentralized and much less linear. Obviously, user roles and workflow models require serious reconsideration when it comes to collaborative content creation.

In order to understand the peculiarities and special requirements of collaborative content creation better, we have studied three popular collaborative content creation systems – Wikipedia (http://www.wikipedia.org), Google Knol (http://knol.google.com) and Slashdot (http://slashdot.org). Our particular focus was on quality assurance policy, workflow models and the 'user ecosystem'. We have also studied literature containing analysis of popular collaborative content creation systems and proposals for their further development (e.g., Lampe & Resnick 2004; Lih 2004; Anthony, Smith & Williamson 2005; Sanger 2005; Cross 2006; Riehle 2006; Priedhorsky et al. 2007; Wilkinson & Huberman 2007; Adler et al. 2008; Halim, Yongzheng & Yap 2009; Monaci 2009; Welser et al 2008; Wöhner & Peters 2009). This study allowed us to make a detailed comparison between traditional and collaborative content creation. The report itself is omitted from this guide for considerations of space, but we provide a summary of differences between traditional and collaborative content creation in the table below.

Characteristic	Traditional content creation	Collaborative content creation
Nature of work	The work is usually confined within organizational boundaries	The work is not necessarily confined within organizational boundaries. Typically, it is dispersed and decentralized
	Usually project-driven work with a limited time scale	Continuous work
	The work is normally done in accordance with an agreed plan and/or generally accepted methodology	An acceptable common plan is hard to work out and commitment to it may vary
	The work is done on a regular basis	The work is done sporadically
	The work is rather slow but the progress is steady	Speed of content creation may vary a lot but is typically higher than in the traditional model
	The work is rather expensive	*Direct* costs in respect of content volume are usually quite low
Number of participants	Limited number of participants, the minimum being one person	Two or more participants, potentially an unlimited number. Usually large-scale collaboration
Participants' roles and hierarchy	Usually fixed roles throughout the project	Roles may change throughout the content creation process, e.g., by means of voting.
	A clear hierarchy of roles mostly based on content creation tasks	Usually a simpler hierarchy of roles based on content creation tasks, and a sophisticated hierarchy of technical (administrative) roles

	Editor-in-chief normally makes the final decision about the content	There is usually no person who makes the final decision about the content. This may lead to 'edit wars'
	Users cannot edit contributions of other users with equal rights directly, without prior consent	Users can edit contributions of other users with equal rights directly, without their prior consent
Participants' identity	Participants' identity is known and verifiable	Participants can be anonymous or hide behind nicknames. Their identity is rarely verifiable which may lead to identity fraud and vandalism
	Normally most of the participants know each other personally	Participants do not necessarily know each other personally
Participants' motivation	Normally participants work for a fee	The motivation usually includes both individualistic motives (authority, reputation, influence, respect, self-fulfilment, personal achievement) as well as altruistic ones
Working environment	Work can be done by means of personal or virtual meetings and correspondence	Users mostly work in a virtual environment
	Work can be done in offline or online mode	Work is primarily done in online mode
Quality issues	Content is usually based on reliable and verifiable sources.	Content is user-originated or borrowed from sources the reliability of which is estimated by users themselves

	Final quality assurance is usually the responsibility of professional editors	Quality assurance is dependent on collaborative validation
	Quality of content is usually even throughout the product	Articles that cover subjects of intense public interest are frequently viewed and edited by large numbers of people, which promotes their quality. Other articles may not be viewed or edited for a long time
	Draft content is not available to the general public	Draft content is open to the general public
	Updating is cyclic	Updating is continuous
	Content is usually perceived by users as credible because contributions are made by a team of verifiable professionals	Participation of anonymous contributors with varying and unverifiable backgrounds means that content may be perceived as less credible
Freedom of expression and its restriction	Political correctness and ideology issues are dealt with in a centralized way (e.g., by editors or by editors and domain specialists)	Freedom of speech may cause a lot of discussion among users with different ideological backgrounds and views on political correctness
	No need for censorship or censorship is done by professional editors	Self-censorship and moderation are of great importance

Table 1. Comparison of traditional and collaborative content creation.

The differences described above can be *mutatis mutandis* projected onto terminology work. Several specific differences between centralized and collaborative terminology work are listed below.

Characteristic	Traditional terminology work	Collaborative terminology work
Participants and their roles	Work is normally done by professional terminologists in cooperation with domain experts	Work may be done by laymen who are not familiar with the principles of terminology work and terminography
Work methods	Work is usually based on concept analysis. This allows for systematic description and even coverage of terminology	Work is not necessarily based on concept analysis, which may result in unsystematic description and uneven coverage of terminology
Languages	The number of languages is limited	A greater, potentially unlimited number of languages
	Usually a clear distinction between the source language and target language(s)	All languages are viewed as equal, so distinction between the source language and target language(s) is not necessarily made

Table 2. Comparison of traditional and collaborative terminology work.

The following features of collaborative terminology work seem to be the essential ones:

1. **Number of participants**. The concept of collaborativeness implies that the work is done by at least two people.
2. **Working environment**. Collaborative work in its modern sense is always done in an electronic environment.
3. **Direct editing without prior consent**. Users can edit contributions of other users directly and without prior notice.

Based on these characteristics, we can define collaborative terminology work for the purposes of this guide in the following way:

terminology work conducted in an electronic environment in such a way that users can edit contributions of other users directly and without prior notice.

3.3.3 Benefits and challenges of collaborative terminology work

Like any other way of doing terminology work, the collaborative approach has its strengths and weaknesses. In this section, we envisage the benefits and challenges of collaborative terminology work and propose counter-measures against the expected risks.

Benefits of collaborative terminology work

Based on the experience obtained from the use of modern collaborative content creation systems, the following benefits of the collaborative approach for terminology work are envisaged:

- the ability to communicate across space and time through the use of an electronic environment
- better prerequisites for keeping content up-to-date due to the continuous character of work
- more objective picture of the domain due to the involvement of various stakeholders
- a good chance to preserve and activate 'silent knowledge' and 'grass-roots knowledge' in organizations
- usually a higher speed of content creation in comparison to the traditional methodology
- usually lower direct costs of content creation in respect of content volume.

We would like to stress that these benefits are potential and do not apply automatically as soon as one starts using a platform for collaborative terminology work. Quality assurance infrastructure and especially methodological guidelines in particular are one of the key factors in unlocking the potential of the 'wisdom of the crowd'.

Challenges of collaborative terminology work and their management

Collaborative terminology work not only provides opportunities but also involves many challenges. The most obvious problems and possible ways of managing them are described below. When several solutions are proposed, they usually supplement rather than exclude each other.

1. Problem: Lack of skill and experience in terminology work and terminology management among the participants.

Possible solutions:

A. Help, guidance and technical support that can be divided into three major categories:

- Introductory training
  - Online or contact training, a video record of which is available for future reference.
  - Tutorial on the principles of systematic terminology work based on a small sample glossary of terms from a domain of general interest. Wine ontology in the *Ontology Development 101* tutorial (Noy & MacGuinness 2001) and pizza ontology in the ontology editor *Protégé* (http://protege.stanford.edu) may serve as examples of successful (although somewhat outdated) tutorials.
  - Welcome tutorial on the use of the terminology management system ('Getting started with…').
  - Tip of the day.
- Stationary help
  - *Help* section.
  - *FAQ* section.
  - Glossary.

- Interactive help
  - Context help.
  - Wizards.
  - Collaborative help instruments: *Help* forum, *Help* chat, etc.
  - Help desk.

Multilingual versions of the tutorial and other help documents may be produced collaboratively by the community itself.

B. System of awards and other incentives for completing the training and giving help to others.

2. Problem: Lack of motivation to do collaborative terminology work in organizations and companies.

Possible solutions:
- New and potential users should be informed about the benefits of terminology work and terminology management, with references to authoritative studies on the subject (such as Tekom Study 2010, testimonials, 'stories of success', etc.
- Intuitive and user-friendly interface, help and guidance as well as a library of ready models and templates for different types of terminology work can lower the barrier to start using the system.
- A motivating reputation system.
- A system of awards (e.g., badges for high-quality contributions).
- Automatic e-mail reminders on no-show for a certain period of time.
- Occasional incentives aimed at stimulating user activity.
- Monetary incentives for contributors (based on high quality rather than on quantity of contributions).

3. Problem: User identity is usually not verifiable which may lead to identity fraud and vandalism.

Possible solutions:

- Registration only under one's real name.
- Verification of corporate users by companies.
- Verification of individual user registration by SMS.
- Users' IP-addresses should be registered in the system.
- Automatic alerts to moderators if most of the text in a term record is deleted.
- Persistent *Edit history*.

4. Problem: No general model for creation of terminological collections can be provided owing to the varying backgrounds and requirements of terminological projects and content creators. This may result in very heterogeneous content and complicate operations on the data, such as search and exchange.

Possible solutions:

- Classification and examples of LSP designations that can be objects of terminological description in a term bank.
- Elaboration of minimum requirements for terminological descriptions and related metadata.
- Reminders about the need to fill in the fields that are important from the point of view of quality assurance (e.g., definitions and source labels).
- Elaboration of the core domain classification.
- Standard-based selector of language and country codes.
- Templates for the documentation of sources.
- Templates and models of terminological records and collections.
- Sample plan or 'roadmap' for starting a new terminological collection.
- Incentives for users who contribute to the creation of models and templates.

5. Problem: A large number of languages with no clear source and target language(s) multiplied by combining several units of description which represent slightly different concepts may result in overloaded and messy entries.

Possible solutions:

- It should be possible to customize views in such a way that only those languages and types of information are visible in which the user is currently interested.

- The primary object of terminological description should be a *terminological lexeme* (a set of forms of an LSP designation sharing a common meaning – see *Section 3.2.6*).

- During editing, one should only be able to add information related to the topic of the record ('headword') but not interrelated designations.

- The default 'browse view' should display only one terminological lexeme and its description. Full description of related designations (synonyms, equivalents, etc.) may be opened in a new window/tab/pane. Partial description of interrelated designations (without cross-language equivalents) may be opened as an insert in the current view.

- Other types of views (concept-oriented, term-oriented, etc.) should be available on demand.

6. Problem: Lack of reliability and credibility of terminological content created collaboratively.

Possible solutions:

- Detailed user profiles with information about the linguistic, domain and terminological expertise of each user (see *Appendix 5*).

- Registration only under one's real name. The risk of authentication fraud can be minimized by requesting information that helps identify a person, such as a mobile phone number for individual users and a corporate e-mail address for corporate users.

- A user reputation system based on trust by the community and trust by the platform's administrators.

- Verification of data by trusted contributors.

- Indication of sources and their proper documentation.

- Collaborative rating of sources.
- Quality and reliability rating at the collection, record and data field level.
- The opportunity to subscribe to modifications made to term records of one's domain of interest (or certain language sections in them).
- Recruiting of authoritative domain experts, linguists, editors and proofreaders for revision of distinguished terminological collections.
- Enhancing the general quality of collaborative terminology work by training the contributors, providing help and technical support.
- Enhancing the systematic character of collaborative terminology work by providing guidelines, models and templates.
- Promotion of concept analysis during the training and by providing tools for manual and automatic creation of concept charts.

7. Problem: There is no person who makes the final decision about the content. This may lead to unnecessary deletions and 'edit wars'.

Possible solutions:
- Request for comments/votes from other users to help resolve the dispute if an edit war has started.
- *Signal to moderator* button on the discussion page of each term record.
- Limitation of reverts per user for a given period of time.

8. Problem: Freedom of speech can cause controversy between users with different backgrounds.

Possible solutions:
- Representation of all significant views fairly, proportionately and without bias (cf. the 'neutral point of view' principle in Wikipedia).

9. Problem: Users may intentionally or unintentionally violate copyright laws or other national or international legislation.

Possible solutions:

- Upon registration, new users should accept the terms of the licence agreement stipulating among other things that it is the responsibility of users to make sure they do not violate copyright or other laws by publishing materials on the system.
- The importance of copyright and legal compliance of the published data should be stressed in the rules, tutorials, etc.
- Reporting copyright abuse to moderators should be made easy.
- Users should be prompted to carefully document sources and borrow text only in such volumes that qualify as citation.
- Modifications made to the contents borrowed from other sources should also be indicated.
- Indication of sources at every level (collection, record, data field, individual string) should be made easy.

10. Problem: Copyright to terminological content created collaboratively.

Possible solutions:

- To prevent disputes over authorship, users should not have copyright on collaboratively created content.
- Commercial use of collaboratively created content in public collections should not be allowed.

3.3.4   Recommendations on the organization of collaborative terminology work on the *TermFactory* platform

Basic principles

The proposed model is built on the following main principles:

- All the members of the community can edit term records directly and in the real-time mode.

- Modifications to term records are stored in *Edit history* and can be reverted if necessary.

- There are two types of terminological collections: public and private.

- There are two types of trust that need to be counter-balanced: trust by administrators and trust by the community.

- Quality is achieved collaboratively with the help of a reputation system, user moderation, content rating and verification.

Types of tools for collaborative work

The main tool for collaborative terminology work is *TermFactory Wiki*, which enables direct and real-time editing of term records. Previous versions of records are stored in *Edit history* and can be reverted, as in Wikipedia (see http://en.wikipedia.org/wiki/ Help:Reverting, accessed 22.2.2013). No entry locking is used, but a notice is issued if a record is being edited by another user.

Users may communicate with each other with a number of tools, the most important of which are discussion pages associated with term records and *TermFactory Forum*. Users may discuss the description of the term in question on the discussion page related to each term record. *TermFactory Forum* is a regular Internet forum that consists of a number of subforums, such as:

- *General Help on the TermFactory Platform* devoted to technical and organizational problems.
- *General Help on Methodology of Terminology Work* for general methodological questions.
- *TermFactory Content* reserved for discussions on term candidates and other issues that are not directly related to any particular term record.

Public and private collections

Terminological collections in the *TermFactory* platform are divided into two categories: the *Public Collection* which is open to anyone, and private collections which are open to the members of the corresponding communities. Private collections can be created by

- an individual or a group of individuals
- an organization or a group of organizations
- organizations and individuals together.

The initiator of creation of a private collection becomes its administrator and can invite other users and delegate rights to them. If the administrator invites other users, he or she may no longer terminate the collection without their consent. If the administrator wants to terminate the collection, he or she sends a request to the platform's administrator who issues an announcement to all the members of the corresponding community. If no objections are registered in a reasonable amount of time, the collection is terminated.

In a private collection, records can be created from scratch or copied from the *Public Collection*. Private collections can be merged into the *Public Collection* by the administrator of the private collection with the consent of the community of the closed collection.

Registration and authentication

Direct editing without prior moderation presupposes strict security procedures. In particular, registration under one's real name must be compulsory. During the registration users have to provide information about their expertise in terminology work, multilingual communication and domain expertise (see *Appendix 5*). The IP addresses of users should be registered in the system to help combat abuse.

For users who register as individuals and not as corporate users, registration should be verified by an SMS sent to the user's mobile phone. For corporate users, the organization's system administrator provides the platform administrator with logins (in FirstName.LastName form) and the corporate e-mail addresses of the employees, who then receive an activation code via e-mail. After activating the account and filling in the profile, corporate users are registered with the system and their name is marked with a 'verified corporate user' or 'verified by… (company name)' label.

Registering only for a private collection is not allowed. The user account is the same for all collections but the profile can be customized, e.g., contact information may be made visible in a private collection and hidden in the public one.

Moderators

Supervisor moderation can be used as protection against behavioural problems such as insulting posts, personal attacks, spam and other kinds of abuse. If contributors notice problems of this kind, they can use the *Signal to moderator* button and report the problem. As the content is multilingual, there should be moderators for various languages. Moderators can be appointed by the administrator of the platform (for the *Public Collection*) or administrator of a private collection based on user requests for

moderatorship. A good set of criteria for selecting moderators can be found in Wikipedia's 'Guide to Requests for Adminship' (http://en.wikipedia.org/wiki/Wikipedia: Guide_to_requests_for_adminship, accessed 27.11.2012).

Reputation system

The reputation system proposed for the *TermFactory Wiki* was inspired by *Slashdot* (http://slashdot.org), one of the oldest and most popular collaborative weblogs in existence. Wikipedia praises it as a definitive example of a user moderation system in the context of Internet forums (http://en.wikipedia.org/wiki/Moderation_system, accessed 27.11.2012). Slashdot's highly sophisticated reputation system was adapted and modified for the needs of collaborative terminology work. In *TermFactory*, the reputation system is supposed to be based on two kinds of trust: trust by the community and trust by administrators.

Trust by the community

Trust by the community means that users rate each other's contributions and assign positive or negative *karma points*. Each contributor gets a certain number of points on a regular basis (for example, once a week). The higher the reputation level of a user, the more karma points he or she gets. Karma points do not expire but new points are not distributed until the old ones have been used. Additional karma points which are valid for a short period of time may be given to users every now and then in order to offer them an extra stimulus to log into the system and rate contributions.

Users give their karma points to other contributors for textual contributions (e.g., foreign equivalents, definitions and notes) they have particularly liked or disliked. Contributors are encouraged to focus more on rating good work: 70% of the distributed points have to be positive.

Karma points can only be assigned to contributors by rating their contributions, not directly. Contributions can be rated both in the actual version of the entry and in the *Edit history*. Users can rate the same contribution only once and only with one point. Karma points are anonymous, i.e., contributors will not know who has rated their work.

Karma points are domain- and language-specific. As each term is linked to some domain and each data field is linked to some language, the system can automatically keep track of how many points the user has acquired in each domain and each language. In this way, each karma point assigned to the user is divided into two halves – domain-specific and language-specific karma. This information is later used for determining the domains and languages in which the user can become a verifier.

Having gained enough karma points, contributors are promoted to the next reputation level, of which there may be as many as ten. Each new level increases the authority of contributors as they are awarded more karma points to distribute. At the same time, the higher the reputation level, the more points are required for an upgrade to the next level. *Table 3* provides an example of how the upgrade scheme may look. The numbers in this table are indicative and their amount has to be adjusted and calibrated for the purposes of each new wiki project.

Reputation level	Number of positive karma points required for upgrade	Number of karma points for distribution
1	+20	5
2	+30	6
3	+40	7
4	+50	8
5	+60	9
6	+70	10
7	+80	11
8	+150	12
9	+200	14
10	Total number of positive karma points = 700	15

Table 3. An indicative example of the 'reputation ladder'.

Having reached the highest reputation level, contributors get the status of *distinguished contributor*, a status which allows them to verify term articles in domains and languages in which they have gained enough karma points.

Additional means of enhancing the motivation of contributors may include virtual 'badges' as well as monetary incentives (in organizations) for high-quality contributions.

Trust by the administrator

Professional terminologists and other contributors who have sufficient experience in terminology work can be verified by the administrator and achieve the status of *authorized contributors* with the highest reputation level. Authorized contributors can verify term records according to their declared language and domain proficiency.

Authorized contributors and distinguished contributors have badges of different colours so that members of the community can decide for themselves which kind of verifier they trust more.

Demotion

Contributors who get negative karma can be demoted to lower reputation levels. In the case of demotion from the highest reputation level, contributors lose their right to verify data. If they are unable to restore their reputation level within a specified period of time (e.g., a month), verifications done by them are deleted from the system or marked as outdated.

Reputation system in private collections

In closed collections, the reputation system functions in the same way as in the *Public Collection* with two exceptions:
- The administrator may customize the number of karma points for the purposes of the closed collection.
- The administrator may choose not to use the 'karma model' but to manually assign the status of *authorized contributor* to selected users.

Karma points gained in closed collections do not count in the *Public Collection*.

Content rating

While the reputation system aims at rating contributors through their contributions, the purpose of content rating is to assess the results of collaborative terminology work by ranging different versions of the same field.

Users can rate content by assigning one to five stars to data fields in the actual version of a record or its previous versions in the *Edit history*. It should be possible to assemble a term record from the versions of fields with the highest rating.

Users can rate an unlimited number of fields. They can change their rating for a particular version of a field but they cannot rate it multiple times. Rating one's own contributions is not allowed.

Data verification

Distinguished and authorized contributors can verify individual data fields in terminological descriptions. Distinguished contributors verify data according to their language- and domain-specific karma, i.e., only in those domains and languages in which they have gained enough positive karma. Authorized contributors verify data according to their declared language and domain proficiency.

Verification can be of two types: standard and quick. Standard verification means that the verifier has used the best systematic terminology work practices (e.g., concept analysis) when verifying term records. Quick verification is used in cases where a limited number of terms which do not necessarily form a concept field have to be verified quickly by an experienced member of the community. Standard and quick verification are marked differently.

Verified data is only marked as verified, not copied to a separate collection. Even if a verified version is later overwritten by other contributions, it can still be found in the *Edit history*.

At least the following data should be provided in the *verification ticket*: type of verification (quick or standard), verifier, verifier's organization (if applicable), verification date and verifier's comments.

If a private collection is exported into the *Public Collection*, verified data fields from the public collection can preserve their verified status only if the verifier of the

private collection has corresponding authority in the *Public Collection*. Otherwise, verification tickets from the private collection are marked as invalid.

Model for professional terminology work

By professional terminology work, we mean systematic terminology work based on concept analysis, which is carried out by a competent team of domain experts, terminologists and possibly other language professionals (editors, language revisers, etc.). Professional terminology work, which is supposed to be carried out in private collections, can be descriptive or normative in nature.

We suggest that for professional terminology work record locking should be used, so that users can work on records in turn. At the beginning of the project, members of the team should agree on the general workflow model. The administrator of a private collection also distributes the roles, such as 'terminologist', 'domain expert', 'editor', 'language reviser', and specifies the languages and domains to which they apply.

For example, it may be agreed that terminologists extract terms and context definitions from texts and then send the material to domain experts. After several rounds of discussion between terminologists and domain experts, data is sent to the editor, and so on. In this way, at the first stage only terminologists can edit terminological descriptions. When they want to contact domain experts, they 'send' the record or broader selection to them by choosing the addressee from the list of users. The domain expert will get a message that his or her attention is needed. Once the domain expert starts working on the records, they are locked from editing by others.

3.2 Objects of terminological description in a term bank

Consensus about what may constitute an object of terminological description in a term bank is a very important prerequisite for successful terminology work. As their name suggests, term banks are supposed to deal with terminology. However, the users' notion of terminology may vary a lot (see Kudashev 2007: 94; Pasanen 2009: 68–139). In terminology science, too, there are different approaches to the concept.

For the purposes of this manual, we prefer to use the umbrella concept *LSP designation* for all kinds of designations of special domain concepts. We fully agree with Gerd (1986: 36–37) who writes that for a terminographer, the actual question is where and how a given designation of a specialized concept should be presented in a terminological reference product rather than whether it should be called a term or not.

In this chapter, we discuss the requirements that LSP designations have to meet in order to be included in a term bank. We provide a classification and examples of expressions that are both likely and unlikely to be objects of terminological description. At the end of the chapter, we discuss various forms in which objects of description may be stored in a term bank, such as lemmatized, normalized and optimized forms.

3.2.1   General requirements for objects of description in a term bank

A term bank is a lexical reference product that provides terminological description of various LSP designations. In lexical reference products, objects of description have to meet certain requirements (cf. Kudashev 2007: 36–43).

The first requirement is that objects of description have to be a part of the lexical system of a given language. This implies the relative stability of the form and meaning of the objects as well as their repeatability in discourse. Statements, texts and other units that are not a part of the lexical system may not be objects of

description in lexicographic reference products. Occasional, contextual meanings of designations are not covered by lexicographic reference products either.

The second requirement is that it should be possible to include objects of description as headwords and to search for them. At the current stage of development of computer technologies, this means that objects of description may only contain alphanumeric characters and symbols supported by Unicode. If the contents of a reference product are supposed to be presented in a linear way, headword-by-headword or entry-by-entry, the objects of description should also be sorted in a predictable way.

The third requirement is that the nature of the object described should allow us to provide it with an adequate terminological description. Pragmatically, there should also be a practical need for such a description. For example, chemical formulae may be used interchangeably with the corresponding verbal designations, but there is not much to say about them in a term bank apart from that. Such designations may be provided as a part of terminological description, as reference articles or index items but there is usually no sense in treating them as full-fledged headwords.

These requirements will be our guidelines in assessing the ability of various LSP designations to be objects of description. In the sections to follow, we provide a multi-facet classification and examples of objects that can be included in a term bank.

3.2.2 LSP designations that can be objects of description in a term bank

The proposed classification is strictly goal-oriented, as we are interested only in those divisions that are likely to influence the inclusion of LSP designations in a term bank. As a starting-point for our classification, we use the list provided under the category 'term type' in ISO 12620: 1999 as well as detailed classifications of LSP designations presented in recent manuals on terminology (e.g., Grinev-Grinevič 2008: 59–66; Leičik 2009; Šelov & Leičik 2012: 12–32). We classify LSP

designations according to several facets reflecting the semantic, morphosyntactic, and pragmatic levels.

Lexical LSP designations according to the type of concept

Terms

According to the latest ISO standard on terminology work, a *term* is a designation consisting of one or more words representing a general concept in a special language in a specific subject field (ISO 704:2009: 34). We consider this definition too broad and suggest that a refinement proposed by Leičik (2006: 32 and earlier works) should be applied, which is that terms denote general concepts that reflect a particular theory and thus serve as elements of relatively well-structured concept systems.

Prototerms

Cases in which concepts in a new special subject field are organized from the very start into a coherent concept system are quite rare. This usually happens at a relatively late stage of the field's development (Kandelaki 1970: 42). *Terminologization* is often a long process, and LSP designations typically go through various stages before they become fully-fledged terms. One such stage is commonly referred to in Russian terminological literature as the *prototerm* stage (e.g., Grinev 1993: 49; Leičik 2006: 77 and earlier works).

The delimiting characteristic of prototerms is that their borders with adjacent concepts are not yet clear-cut. Prototerms denote vague *special notions* in which the essential characteristics of objects and their relations with similar objects are not yet

fully identified (cf. Grinev 1993: 31–32). Because of this, special notions lack a stable place in the concept system and a strict definition.

Prototerms may occur in any domain but they are particularly common in branches of special knowledge that lack a solid theoretical basis. The designation *ghost* from the glossary 'Key Words Frequently Used in Parapsychology' (http://www.parapsych.org/glossary_e_k.html#g, accessed 9.12.2012) may serve as an example of a prototerm: 'this term denotes only the apparition of a deceased person, and is not sufficiently precise for use in psychical research'.

It should be noted however that the lack of general acknowledgement of a particular phenomenon is not a reason for considering it a prototerm. For example, many theoretical concepts in quantum physics are perfectly well defined and classified, although their existence has not yet been proved experimentally. The absence of a clear-cut verbal definition is not a reason for automatically considering an LSP designation a prototerm either, as the corresponding concept may be defined contextually or with the help of non-verbal means and special notations.

Prototerms should not be mixed with *ambisemic terms* (see Tatarinov 1996: 168–174). Ambisemy is variation of meaning resulting from the fact that each scholar tends to interpret terms in a slightly different way from his or her colleagues and predecessors. Concepts also evolve in time, so that the same LSP designation may appear in a slightly different sense even in works by the same author. While special notions denoted by prototerms lack a stable place in the concept system, concepts denoted by ambisemic terms simply belong to slightly different concept systems.

In terminological literature, especially in Russian sources (e.g., Grinev 1993; Leičik 2009; Šelov & Leičik 2012), one can come across the names of many other classes of LSP designations that are undergoing various stages of terminologization, such as *preterms*, *terminoids*, *quasiterms*, or *pseudoterms*. This topic obviously requires more research, as the existing classifications have been criticised as somewhat fragmentary and unclear (e.g., Tatarinov 1996: 260). Detailed classifications of 'preterminological' LSP designations is more likely to be of interest

in academic terminological projects (e.g., historical ones) while most working terminologists will probably find them too complicated and irrelevant for their purposes. For these reasons, we won't go further into this subject.

To our mind, prototerms and other LSP designations denoting vague general concepts fulfil the minimum criteria for inclusion in a term bank, especially a collaborative one. In fact, a public discussion on the intention of vague concepts and their relations with related concepts may help prototerms to become full-fledged terms.

Appellations

Appellations denote individual concepts, i.e., concepts that correspond to only one object, *Andromeda Galaxy* and *Large Hadron Collider* being examples. Special domain appellations are rarely included in term banks but there are hardly any theoretical or pragmatical obstacles to doing so. Users do not seem to mind either. For example, in a survey conducted by the Swedish National Term Bank (Rikstermbanken) in 2011, over 31% of users mentioned appellations when answering the question 'What content is lacking from the current term bank?' (Nilsson 2012).

Nomenclature

Nomenclature is usually defined in the western terminological tradition as terminology structured systematically according to pre-established rules (e.g., ISO 1087-1: 2000). Chemical and biological nomenclatures are often called typical examples of nomenclature. Since the degree of 'pre-establishedness' of the naming rules may vary a lot, it is mostly a matter of tradition to call certain terminological systems nomenclature. Interestingly enough, some 'classical' nomenclatures have recently

changed their names (e.g., *Nomina Anatomica* has become *Terminologia Anatomica*).

In the Soviet/Russian terminological school, nomenclature is comprehended in a different way. The focus is not on the pre-established naming rules but on other characteristics. Since early works by V. Vinokur and A. Reformatskij dating back to the 1930s and 1950s, nomenclature has been opposed to terminology as a system of designnantions denoting groups of objects rather than concepts. As for now, there is still no single opinion on the concept of nomenclature, but most Russian authors seem to agree that names of objects of mass production (such as 'Dreamland Soft' mattress, 'Delux Beauty Relax' pillow or 'Ecomoods Fabia' lamp) are typical examples of nomenclature (e.g., Reformatskij 1961; Kandelaki 1973: 64; Grinev 1993: 44–45; Šelov & Leičik 2007: 68).

Leičik (1974: 24) called nomenclature an intermediate class between appellations and terms. In the case of appellations, one cannot really talk about full-fledged conceptualization because appellations denote individual concepts, i.e., concepts the extension of which consists of only one object. In the case of terms, the conceptualization level is quite high, as many inessential characteristics are ignored in the process of forming general concepts. Nomenclature lies in between appellations and terms in this respect, as it denotes groups of objects that are perceived as identical or almost identical (e.g., objects of mass production that are produced according to the same model and/or manufacturing method).

One of the important implications of this fact for terminology work is that concepts denoted by names cannot be assigned a concise genus-species definition (cf. Kandelaki 1973: 63; Bereznikova 1976: 88). For example, *revolver* is a term, and it can easily be assigned a genus-species definition, because revolvers have a special principle of operation that distinguishes them from other pistols. At the same time, describing the difference between the *Smith-and-Wesson model 13* and *Smith-and-Wesson model 27* revolvers requires a comparison of their specifications: design, technical characteristics, etc. (cf. Leičik 1974: 20–21). Formulae of invention in patents may give some idea of how extensive and complex such specifications may

be. While general concepts denoted by terms are defined by ignoring most of the essential characteristics and focusing only on the delimiting ones, concepts denoted by nomenclature are described by including as many essential characteristics as possible or reasonable (cf. Šelov & Leičik 2007: 7, with references to the works by an the English scientist and philosopher W. Whewell).

Many companies would gladly include the names of products in the term bank, as there is a need to deal with them in a multilingual environment (e.g., to translate and localize them). The following considerations should be taken into account however (cf. Kudashev 2005):

- Nomenclature outnumbers terminology by hundreds of times, so the term bank may be flooded with nomenclature, which is hardly reasonable.
- Nomenclature and its descriptions become obsolete much faster than terminology and have to be revised and updated much more often.
- Nomenclature should not be included in a term bank if no relevant terminological information can be provided about it. For example, if the name of a model is an alphanumeric sequence that is used globally and does not require localization, there is probably no point in including it in a term bank.
- Organization of origin and other identification information related to nomenclature must be specified clearly and unambiguously.

The scope of nomenclature is not necessarily restricted simply to the names of objects of mass production. For example, the bottom-most level of the biological taxonomy (species) is very similar to objects of mass production in a sense that they cannot be be assigned a concise genus-species definition either. Instead, they are 'defined' with the help of quite extensive scientific diagnoses (Šelov & Leičik 2007: 14) which describe a typical specimen of the given species. Distinguishing one species from another requires various measures, such as similarity of DNA, morphology and ecological niche. In addition to verbal description, registering a new species usually requires a conserved *type*, e.g., dried plant material deposited and preserved in a herbarium (McNeil et al. 2012: Art. 7–10). There is a clear analogy here with models and technical specifications used for describing objects of mass

production, although the intension and extension of the concepts denoted by names of species are probably fuzzier than those denoted by the names of technical artefacts.

Obviously, there are many more domains in which special designations behave like technical nomenclature. Even after decades of disputes on the nature and classification of nomenclature, the problem does not seem to be finally resolved, and further research on the subject is needed. In any case, term bank users should be aware of this group of LSP designations and their specific features, including the fact that they can not be defined in the same concise way as terminology.

LSP designations according to part of speech

Lexical LSP designations may be nouns, verbs or adverbs (cf. Wright 1997: 13), for example:

- a programme
- to programme
- programmatically.

Adjectives and participles (such as programmatic, programmed) do not normally function as independent LSP designations but rather serve as *term elements* (discussed below). Substantivized adjectives and participles are an exception to the general rule.

Verbs and adverbs are rarely included in LSP glossaries although they play a very important role in some domains (e.g., verbs in cookery and adverbs in musicology). One of the reasons for this may be that there is no consistent and generally accepted methodology for defining them. However, this should not be an obstacle to including LSP verbs and adverbs in a term bank.

LSP designations according to their morphological structure

Morphological characteristics of LSP designations may vary between languages. Below is an example of a morphological classification of LSP designations for the English language:

- simple word (word containing only one root), e.g., *sound*, *light*
- complex word (word formed from a simple word by the addition of one or more derivatives), e.g., *accountability*, *partnership*
- phrasal verb (combination of a verb and a preposition, a verb and an adverb, or a verb with both an adverb and a preposition, any of which are a part of the syntax of the sentence, and so are a complete semantic unit), e.g., *take off*, *tank up*
- compound word (word containing two or more roots), e.g., *know-how*, *airstrike*
- abbreviated form (designation formed by omitting any parts from a longer form), e.g., *DNA*, *adj.*, *flu*
- multiword designation (expression consisting of two or more words), e.g., *currency exchange losses*, *direct employment effect*.

Users of a term bank should be aware of the morphological diversity of LSP designations, so that none of the groups is arbitrarily excluded.

LSP designations with usage restrictions

LSP designations may be restricted in use by many factors, the most common of which are the following:

- geographical area, e.g., British English term, American English term, dialect expression
- time period, e.g., obsolete term, neologism, term used during WWII
- register, e.g., official term, informal term, professional slang expression
- organization, e.g., term used by / preferred in Apple, Microsoft, UN, WHO

- proprietary restrictions, e.g., trademark and trade name
- scientific school/theory, e.g., term specific to Einstein's physics, Danish structuralism (in linguistics)
- professional group, e.g., term primarily used by physicians/nurses
- normative regulations, e.g., term that is preferred, recommended or non-recommended by some authoritative body (e.g., standard organization).

There are no obstacles to including LSP designations with usage restrictions in a term bank but users should be prompted to provide them with appropriate usage labels and notes. It is also important to make users realize that it is usually better to include a non-recommended or otherwise restricted term and clearly mark it as such than to omit it. If certain types of designations with usage restrictions are deliberately excluded from the scope of a terminological project, this should be mentioned in the working guidelines and the user guide.

A special case of designations with limited usage is names of classes in various classifications, such as the names of diseases in WHO's *International Classification of Diseases* (ICD, see http://www.who.int/classifications/icd/en, accessed 3.1.2013). We have argued in Kudashev 2012 that names of classes in ICD are not really a part of the conceptual space and lexis of any national medical LSP, for several reasons.

First, the function of ICD, to help keep statistics of diseases and causes of death, is limited and very pragmatic. The ICD aims at mapping national concepts and terms onto international classes rather than replacing them. Second, the form and meaning of ICD classes are 'frozen' and do not evolve until the next revision of the classification. In other words, they are relevant only as long as the corresponding classification is valid. Third, the complexity, length and occasional clumsiness of some names of ICD classes prevent them from becoming lexical units of the medical LSP. Fourth, some names of ICD classes are of special character and do not have counterparts in living medical terminology.

Even if the names of ICD classes formally coincide with medical terms, this should be understood as homonymy between two different systems and areas of application: a dynamic, multifunctional medical LSP and a static, single-function

classification. We would recommend keeping classifications like ICD as separate, read-only resources. Terminological reference products and classifications have different functions and different conventions that should not interfere with each other. At the same time, they can be interlinked. For example, a terminological description may contain a reference to the relevant classes from one or more classifications.

Means of formal notation in LSPs

In LSPs, special concepts may be also referred to by means of formal notation; for example:

- special symbols, e.g., §, €, °, ∑, ∞
- formulae, e.g., $[As@Ni_{12}As_{20}]^{3-}$
- international scientific names, e.g., *Salix starkeana* subsp. *cinerascens*
- code names, e.g., PHARC, ABHD12A, BEM46L2, C20orf22, dJ965G21.2 are code names of a gene with the official full name *abhydrolase domain containing 12*
- catalogue names, e.g., FK5 538, CP(D)−60°5483, GC 19728, CCDM J14396-6050 are catalogue names of the star *Alpha Centauri* (also known as Rigil Kentaurus, Rigil Kent, Toliman and Bungula).

Formal notations are often used in LSP texts interchangeably with the corresponding appellations and terms, and are sometimes the only existing designation of a special object or concept. For example, in astronomy, only about 300 stars and a couple of galaxies have proper names, while others are denoted by various catalogue names (http://en.wikipedia.org/wiki/Astronomical_nomenclature, accessed 15.1.2013).

When there are parallel lexical designations, means of formal notation usually become a part of terminological description rather than its object. However, they may be made searchable, for example, by providing them as the headwords of reference articles.

If there are no parallel lexical designations, the means of formal notation may become objects of description in their own right. However, as with nomenclature, one should consider what kind of terminological information could be provided about them in a term bank. We recommend considering the status of the means of formal notation in each domain and terminological collection individually.

It should be noted that the means of formal notation often requires a reference to the classification or catalogue to which it relates, including the version number. As with classifications mentioned above, it is often reasonable to keep terminological reference products and various catalogues of code names as separate although interlinked resources.

As one can see from the examples above, means of formal notation may be particularly rich in typographic features and contain characters of foreign alphabets, digits, punctuation marks, special symbols as well as inline formatting, such as italics, subscript or superscript. This implies that a terminology management system has to support the whole range of Unicode symbols, allow for marking individual parts of data fields as belonging to a different language, and support the addition and management of inline formatting. These functional features are required not only for dealing with means of formal notation but also for processing conventional terms containing such elements (e.g., β-carotene, *x*-coordinate, CO_2 laser).

Lexicalized units

In addition to lexical units, a term bank may also contain *lexicalized LSP expressions*, single-word or multi-word expressions that have a relatively stable form and function in a particular LSP or special area of application. Below are a few examples of lexicalized LSP expressions:

- instructions, e.g., 'Handle with care', 'This end up' (ISO 12620:1999: 10)
- drill commands, e.g., 'Double time, MARCH!', 'Left shoulder, ARMS!', 'Right, FACE!', 'Rest!'

- set phrases used in radio and signalling: 'More to follow', 'How copy?', 'Solid copy!'

Lexicalized expressions cannot be defined in the same way as lexical designations but can be provided with a description of their function, i.e., situations in which they are used. For example, 'Double time, MARCH!' can be defined as 'a drill command: an order to jog in time'. The same type of 'metalinguistic definition' is used in general lexicography for describing the function of expressions like 'Hello!' (Geeraerts 2003: 87–88).

In some domains, lexicalized expressions play a very important role, and there is every reason to include them in terminological reference resources.

Term elements

Term elements are productive components of terms that have a relatively stable meaning in a given LSP. For example, in medicine

- the prefixes 'a-', 'an-' mean an absence of something (e.g., *a*pathy, *a*nalgia)
- the suffix '-ac' means 'pertaining to something' (e.g., cardi*ac*)
- the root 'aur(i)-' means 'pertaining to the ear' (e.g., *aur*al).

In chemistry, prefixes, suffixes and infixes are used to describe the type and position of functional groups in the compound. For example, (http://en.wikipedia.org/wiki/IUPAC_nomenclature_of_organic_chemistry, accessed 29.12.2010):

Functional group	Formula	Prefix	Suffix
Amines	$-NH_2$	amino-	-amine
Imines	$=NH$	imino-	-imine
Hydrazines	$-NHNH_2$	hydrazino-	-hydrazine

Table 4. Term elements in the domain of chemistry.

In domains like medicine and chemistry, special *keys to terminology* exist (e.g., Vasil'eva 1988; Merritt & Bossenbroek 1997: 233–237) which help decode complex terms or create new ones according to the established rules.

Adjectives and participles as well as elements of compound words can also function like term elements. In *Example 1* below, borrowed from the the Finnish-Russian Forestry Dictionary (Suomalais-venäläinen metsäsanakirja 2008), the adjective *pyöräalustainen* (wheeled, wheel-, on a wheelbase) is presented as a headword in its own right, because practically any forestry machine can have a wheelbase and it is neither possible nor necessary to enumerate them all. A couple of examples at the end of the entry provide a model for translating terms that include this component:

(1) **pyöräalustainen** 📖 Uusitalo, 60 [Koneet ja laitteet]
pyörillä varustetulla alustakoneella sijaitseva
колёсный 📖 Уситало, 62
на колёсном ходу 📖 Жаденов & Заикин, 61
◊ pyöräalustainen kaivinkone – *колёсный экскаватор, экскаватор на колёсном ходу*; pyöräalustainen metsäkone – *лесохозяйственная (лесосечная) машина на колёсном ходу*

According to the translators' survey that we conducted for our PhD study, over 50% of translators support inclusion of term elements in LSP dictionaries and only 16% are against it (Kudashev 2007: 191). When incorporated into a term bank as objects of description in their own right, term elements may be provided with a more comprehensive description of their meaning, etymology and usage, including term formation models and examples of usage.

3.2.3 LSP units that are unlikely to be objects of description in a term bank

A few types of LSP unit mentioned in the *Handbook of Terminology Management* (see Wright 1997: 15–16) and listed in the ISO 12620:1999 data category inventory

under the category 'term type' do not meet the general criteria for LSP designations and thus are unlikely to become objects of terminological description in a term bank.

Equations (e.g., E=mc^2), *logical expressions* (e.g., x ≠ y) and *collocations* (e.g., 'immunization *against* smth.') are neither lexical nor lexicalized units. These units may be *a part of terminological description* (e.g., collocations demonstrate combinatory restrictions of LSP designations), but they can hardly be an *object* of description.

Extensive recurrent chunks of text, such as *standard texts* (e.g., standard force majeure clause), are also very questionable candidates for inclusion in a term bank. A more natural place for storing such passages is a translation memory system.

3.2.4　'Impurities' in term banks

Ideally, a terminological reference resource should only contain lexical or lexicalized LSP designations. However, in practice it is hard to avoid certain 'impurities', the most common of which are LGP designations, combinations of LSP designations, and combinations of LGP and LSP designations.

LGP designations

In practice, distinguishing between LSP and LGP designations is not a trivial task, as common words can be used in the function of terms and vice versa. The *term* concept is also understood differently by different users and groups of users. For example, many translators believe that terminology is 'all the words I don't know', the words that are missing from the dictionary, or the latest jargon (Bonono 2000: 648). Another popular but obviously misguided criterion for distinguishing terms is that they can have only one equivalent in the target language (Miram 2001: 62–63). Experiments conducted by Pasanen (2004: 246; 2009: 68–101) have demonstrated

considerable variation in the term identification process, even when people with similar educational and professional background extract terms from the same text.

This makes it difficult to guarantee that a collaborative term bank is free from LGP designations. Their presence is not a problem as long as the systematic character of terminological description is not disturbed and the users are not disoriented as regards the nature of the reference resource.

Users who do not have deep knowledge of the LSP they are dealing with (e.g., translators) may use a number of indirect indicators for identifying LSP designations. One such indicator is the presence of a strict, 'technical' definition that modifies and/or makes more precise the common interpretation of a general language word or word combination. For example, in the Finnish general language, the word *pitkäaikaistyötön* (long-term unemployed) means a person who has been unemployed for a long time. At the same time, the 'technical' definition used by the social services is more restrictive: a person who has been unemployed *continuously for more than one year*.

Other indirect indicators of LSP desinations include their statistical distribution and special markers, such as quotation marks, italics, boldface print, the expressions 'so-called', 'known as', etc. (e.g., Pasanen 2009: 161–254). Special term extraction software exists that is capable of extracting term candidates from texts with the help of statistical distribution and/or other indirect indicators. Automatic term extractors may be of some help at the initial stages of a terminological project, but for the time being they cannot compete with a team of professional terminologists and authoritative domain experts who rely on their experience, knowledge of the domain and proven methodology, such as concept analysis.

Combinations of LSP designations and combinations of LGP and LSP designations

Another major problem for both manual and automatic extraction of LSP designations from the texts is that it is not always easy to determine whether a particular candidate is a single LSP designation, a combination of two or more LSP designations, or a combination of two or more LSP and LGP designations. Consider the following randomly picked Wikipedia articles (accessed 30.12.2010):

- Anterior triangle of the neck
- Point groups in three dimensions
- Fundamental theorem of arithmetic.

Statistical analysis of a large text corpus may help answer the question of whether these headwords are LSP designations in their own right or combinations of LSP designations, but again, concept analysis and thorough knowledge of the special domains in question are required for the final answer.

Combinations of LSP designations and combinations of LGP and LSP designations are not supposed to be included in term banks, but their presence is hard to avoid, especially at the initial stages of terminological work. The methodology used in systematic terminology work combined with statistical analysis of a large corpus of LSP texts is the best means for keeping such 'impurities' to the minimum.

3.2.5   Forms of LSP designations

In this section, we discuss various forms in which LSP designations may be presented in a terminological reference resource, in particular, LSP designations may be lemmatized, normalized and optimized.

Lemmatized forms

Most lexical units have inflected forms, the number of which may be quite substantial, especially in languages with rich morphology. The traditional way of dealing with inflected forms in hardcopy reference resources is to choose the so-called 'canonical form' ('lemmatized form', 'dictionary form', 'lemma') that represents the whole paradigm. Non-canonical forms are usually ignored.

The electronic environment enables extended support for non-canonical forms. The first, rather laborious approach is to include all non-canonical forms as reference articles that redirect users to the lemmatized form. The second option is to 'outsource' the redirection to lemmatization modules. A combination of the two approaches is also possible: for example, the system may first check whether a non-lemmatized form can be found in the database and invoke the lemmatization module upon failure.

One of the benefits of using a lemmatizer is that it can also cut off unnecessary elements, such as particles and possessive suffixes in agglutinative languages like Finnish. Lemmatizers become more and more widespread, although algorithms and vocabularies developed for the processing of LGP text are not necessarily fully applicable to LSP texts.

Since most lemmatizers are language-specific, correct language indication of both headwords and queries is of critical importance. Lemmatization often requires disambiguation, as the same non-lemmatized form may be related to two or more records (e.g., 'programmes' is both the plural form of the noun 'programme' and the third person form of the verb 'to programme').

The ability of a terminology management system to process non-lemmatized forms has at least two benefits. First, it improves the user-friendliness of the system, because users get more rapid access to terminological records. Instead of manually converting a non-lemmatized form, they can simply copy-and-paste it into the system

or even invoke the system from a third-party application through a plug-in. Second, it enables better interaction with NLP applications.

If providing full support for non-lemmatized forms is not possible, it is advisable to include at least those groups of non-lemmatized forms that may present difficulties for the users. One such group is irregular forms, such as irregular plurals (e.g., tooth-teeth, mouse-mice, genus-genera) or irregular verbs (e.g., freeze-froze-frozen, break-broke-broken). Users may have difficulty figuring out the correct canonical form for them. In some term banks, Canadian *Termium*, for example, such irregular forms are searchable (Hutchenson 2001: 673).

Another problematic group is words with unconventional canonical forms. For example, the canonical form for nouns in most European languages is nominative singular. However, in biological dictionaries, the names of genera are usually given in the plural (e.g., *mammals*, *marsupials*, *canids*). In some lexicographical traditions, names of objects which are grammatically not *pluralia tantum* but which are typically used in the plural (e.g., *headphones*, *skis*, *boots*) are also given in the plural (Berkov 2004: 30). Including both forms (one of them as a reference article) is a safe choice in such cases.

Normalized forms

In the compilation of technical glossaries, there seems to be a temptation to modify the standard form of terms in certain ways; for example, to use capitalization and inversion (Wright 1997: 17–18). However, such modifications increase the risk of misinterpretation and complicate the reusability of terminological data, for example, in NLP applications.

The primary form of storing LSP designations in a terminology management system should be *normalized form*, by which we mean the standard form based on context-independent spelling rules. A mnemonic formula that is applicable here is FUN: the form of LSP designations should be Full, Unambiguous, and Natural.

66

In the subsections to follow, I provide a few examples of typical problems related to the form of terms in terminological reference resources, using two medical classifications as examples. The first classification is the international standard on human anatomic terminology *Terminologia Anatomica* (futher on *TA*), and the second one is *International Statistical Classification of Diseases and Related Health Problems, 10th Revision* (hereafter *ICD-10*) maintained by the World Health Organization. The topic is discussed in more detail in Kudashev 2012.

Omissions

LSP designations are not always provided in full in reference resources. For example, in TA, many terms of the second and lower hierarchical levels are provided in an abridged form, e.g., *Deep nodes* instead of *Deep popliteal lymph nodes (of lower limb)* in the example below:

(2) A13.3.05.001 Lymph nodes of lower limb
A13.3.05.011 Popliteal nodes
A13.3.05.013 Deep nodes

Omissions of any kind should be avoided, as the benefits of space saving are certainly outweighed by the risk of mistakes and information loss.

Abbreviations

Well-known and standard abbreviations (such as *HIV* – human immunodeficiency virus) are acceptable as parts of LSP designations (e.g., *acute HIV infection syndrome*). However, good practice requires that the abbreviation should be expanded either in a synonymous form of the LSP designation or in an additional note. The use of resource-specific abbreviations in LSP designations should be avoided.

Capitalization

Only those elements of LSP designations may be capitalized that would normally be capitalized in the middle of a sentence. Otherwise, capitalization of initial or all letters should be strictly avoided, as users will not have the information that they need to use the words properly when writing (Wright 1997: 17).

One should keep in mind that in LSPs, the spelling rules may exhibit pecularities not used in the general language. For example, in biology, the first component in binominal names of species always starts with a capital letter while the second component always starts with a small letter, even when derived from a proper name (e.g., *Caloplaca obamae*).

Substitutions

In hardcopy reference resources especially, recurrent parts of LSP designations are often replaced by a special mark, for example a dash. However, as *Example 3* from the Russian index of TA demonstrates, such a format is not user-friendly and is not suitable for inflected languages, as users may make mistakes while trying to restore the canonical form of LSP designations:

    (3)   Желудочек левый A12.1.04.001 100
                --клапан аорты A12.1.04.012 100
                ----заслонка венечная левая A12.1.04.014 100
                ------правая A12.1.04.013 100
                -----невенечная A12.1.04.015 100

Inversions

In terminological reference resources, an inverted and permuted order of term components is frequently used (e.g., *meningitis, herpesviral* instead of *herpesviral*

meningitis). Inversion allows clustering of similar terms around the main noun and may save space if combined with replacement of term components with a special mark, as was shown in *Example 3* above. However, there is often no way for the user to know whether the inversion is resource-specific or whether the term is actually used in this way in the discourse. Besides, inverted forms complicate the search as well as interaction with NLP applications.

Clustering

No clustering of LSP designations should be used in a terminological reference resource. Each LSP designation should be logically and visually separated from other designations. For instance, in *Example 4* from ICD-10, a layman cannot tell for sure if the term in square brackets (*Hansen's disease*) is a synonym of *leprosy* or *arthritis in leprosy*. This ambiguity has led to a translation mistake in the Russian version of ICD-10.

(4)   Arthritis in leprosy [Hansen's disease] (A30.-+)

Foreign elements

LSP designations may not include any foreign elements, such as notes or parts of codes. For example, parts of codes, such as *(E10–E14 with common fourth character .3)* and *(C00-D48+)* in *Example 5*, have migrated into otherwise well-structured electronic versions of the Finnish and Russian translations of ICD-10 as parts of terms.

(5)   Diabetic cataract (E10–E14 with common fourth character .3)
        Arthropathy in neoplastic disease (C00-D48+)

Optimized forms

As was suggested above, normalized forms should be the primary form of storing LSP designations in a term bank. However, the form of LSP designations may also be optimized for various purposes. For example, in entries, LSP designations may contain elements of inline formatting (e.g., italics or subscript) and even include other data categories, such as hyphenation or accent marks (e.g., CO_2 la-ser). For the purposes of searching, the headwords usually have to be cleared from all extra elements and inline formatting (CO2 laser). In the permuted index, the order of term components is permuted, and recurrent parts of terms may be replaced by a tilde or some other mark (laser: $CO_2 \sim$), and so on.

Storing at least those optimized forms that are hard to generate automatically alongside normalized ones allows one to combine the robustness of a plain-text database with the richness and flexibility of data presentation characteristic of advanced dictionary writing systems. We have found this approach useful in designing our in-house terminology management system MyTerMS.

3.2.6   Terminological lexeme

In LSPs, the same set of forms may be associated with two or more meanings (homonymy), and the same meaning may be associated with several sets of forms (synonymy). As was mentioned above, forms and meanings can be described independently of each other to some extent, but the primary object of terminological description is LSP designation, which is a union of exactly one set of forms with exactly one meaning. Terminological description of an LSP designation is not equivalent to the sum of description of its form and meaning, in the same way as a zygote is not equivalent to the sum of male and female gametes.

In general lexicography, there is a concept of the *lexeme* that denotes precisely this kind of union. ISO 24613 defines a lexeme as an 'abstract unit generally

associated with a set of forms sharing a common meaning' (ISO 24613:2008: 4). This definition would be entirely appropriate for our needs unless lexemes were associated in general lexicography only with words and word-like units.

We suggest that the lexicographical interpretation of lexeme (a set of forms of a word sharing a common meaning) should be called *lexicographical lexeme* (or *single-word lexeme*) while in terminography the concept of *terminological lexeme* may be adopted that would include all forms of LSP designations, including word combinations and lexicalized expressions. The concept of lexeme may thus be broadened to embrace both lexicographical and terminological lexemes. For multiword LSP designations, the *component* metaclass may be used as the connecting link between the terminological and lexicographical lexeme (cf. ISO 24613:2008: 13).

Forms associated with a lexeme may be divided into lemmatized and inflected on the one hand, and into normalized and optimized on the other. The suggested metamodel is summarized graphically in *Figure 1*:
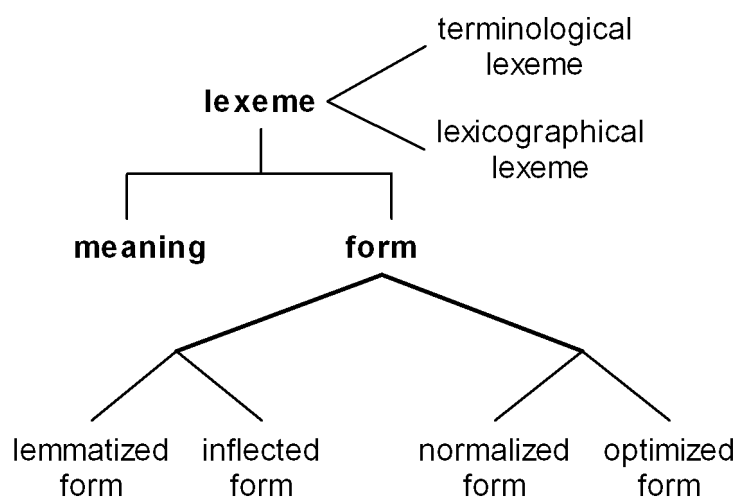

Figure 1 – Modified lexeme and various types of forms.

4 Structural aspects of quality assurance in terminology management

Structural metadata comprises semantic, syntactic and value domain specifications of data classes. In this chapter, we discuss such important types of structural metadata in a term bank as language indication, encoding, collation, domain classification and data category classification.

4.1 Language identification and indication

Indication of the language of textual data is important for searching, data exchange, data reusability and many types of automatic data processing, such as spell-checking, computer-synthesized speech or high-quality print renderings (cf. Phillips & Davis 2009: 3). In a termbank, indication of the language should be mandatory at least for searchable fields and is highly recommended elsewhere.

Terminological descriptions of LSP designations may by default be considered to be in the same language as the LSP designation itself unless the default value is manually overidden. Possible exceptions at the collection level may be negotiated with the administrator of the term bank of the collection in question.

As was suggested in the previous chapter, users should also be able to override the default language of data fields for individual *portions* of text (e.g., the Latin component *de minimis* in *de minimis rule*). Among other things, this increases the speed and accuracy of spell-checking and other NLP tasks.

Language identification is not applicable to data fields containing non-textual data, such as timestamps, counters, Boolean values, or tokens. It may also be difficult to assign a language label to some means of formal notation because of their international character. A standard 'language code' that may be applicable here is *und* (cf. ISO 639-2:1998: 4). In ISO 639-2, it stands for 'undetermined', but it could probably stand just as well for 'undefined'.

The concept of *language* is complex and multidimentional. For example, Phillips & Davis 2009 distinguish the following 'layers':

- primary language, e.g., *en, de*
- region, e.g., *en-GB*, *en-US*
- variant (well-recognised variation of a language, e.g., a dialect)
- script (writing system), e.g., *sr-Cyrl* (Serbian written with Cyrillic)
- orthography, e.g., *de-1996* (German as written using the spelling reform beginning in 1996)
- extended options, e.g., *zh-Latn-x-pinyin* (Chinese transcribed in the Latin script using the Pinyin system).

Yet another relevant dimension is time period (cf. Burnard & Bauman 2012: lv), but there are no standard means for its representation.

For the purposes of terminology management, the first two layers (primary language code possibly supplemented by a country code) are sufficient in the vast majority of cases. As different systems and collections may use a variable number of language indication 'layers', we would strongly recommend keeping them all as separate data categories. They can be merged on the fly if necessary for presentation purposes.

Regional and chronological divisions may be documented in the fields related to the usage restrictions of LSP designations. Script and orthography may be considered as information on the written form of LSP designations.

Codes for the representation of names of primary languages can be found in ISO 639-1 (2002), and country codes in ISO 3166-1 (2006). The names of scripts should be based on ISO 15924 (2004). A short note on the syntactical conventions:

- In ISO 639-1:2002, it is recommended that *language codes* should be written in lowercase (e.g., 'mn' for Mongolian).
- In ISO 3166-1:2006, it is recommended that *country codes* should be capitalized (e.g., 'MN' for Mongolia).

- In ISO 15924:2004, it is recommended that *script codes* should be written in lower case except the initial letter that should be captalized (e.g., 'Cyrl' for Cyrillic).

Indicating the primary language using national flags cannot be recommended. It may prove problematic when a language is used in two or more countries (consider such languages as English, French, or Spanish). Developers of terminology management software have also noted this. For example, in early versions of Trados/SDL *MultiTerm,* national flags were the only available option for indicating the language, but in recent versions, users are given the freedom of choice in this respect.

Use of negation, when a country code is preceded by the NOT operator (e.g., NOT AUS – example borrowed from ISO 12620:1999: 13) is a compact way of specifying regional usage exceptions. However, not all terminology management system and NLP applications are able to interpret such negation correctly.

To avoid mistakes, users should not be allowed to insert language and country labels manually, but should use a special selector with search capability. As standard-based language and country codes are not always transparent, their full versions should be made available to users on demand. For example, the full name of the language or country may appear as a pop-up when the mouse pointer hovers over the code. To speed up the process of adding language and country codes, the system may allow storage of 'favourite' codes in user profiles as well as specifying the 'preferred language code' and/or the 'preferred country code' that will be readily selected in each newly-created term record.

4.2 Character encoding

Character encoding can raise surprisingly complex issues in the management of lexical data (see Burnard & Bauman 2012: lvi-lxviii for a good overview). In order to prevent these, it is important to choose and document the encoding of terminological data in term banks carefully.

Today, the de facto standard of the computer industry is *Unicode* (see http://www.unicode.org/standard/standard.html, accessed 19.1.2013) which aims at consistent encoding, representation and handling of text expressed in most of the world's writing systems. The latest version of Unicode (6.2.0 as of January 2013) consists of a repertoire of more than 110,000 characters covering 100 scripts.

Unicode can be implemented by various character encodings. The most commonly used encodings are UTF-8 and UTF-16. Both encodings represent Unicode's 32-bit code points in an economical way, which accounts for their general use and special place in many standards; for example, in TEI P5 (Burnard & Bauman 2012) and TBX (ISO 30042:2008).

UTF-8 is particularly byte-efficient in the case of the Latin, Cyrillic and Hebrew scripts. However, Latin letters with diacritics and characters from other alphabetic scripts typically take two bytes in UTF-8 but only one byte in UTF-16. Characters U+0800 and above (for example, East Asian scripts) generally take three bytes per character in UTF-8 but only two in UTF-16.

If there were no other considerations than byte-efficiency, the more efficient of the encodings could be selected automatically for each language and script. However, byte-efficiency is only one side of the story. UTF-8 currently enjoys much better support in various APIs and, unlike UTF-16, it does not have issues related to the hardware used (see Burnard & Bauman 2012: lxvi). Besides, the use of only one rather than two encodings makes database design and operations easier, faster and more reliable. We thus recommend that only UTF-8 should be used in multilingual terminology management systems.

With both UTF-8 and UTF-16, care should be taken to prevent possible data corruption due to incorrect conversion; for example, during copy-and-paste operations in the user interface (see Burnard & Bauman 2012: lxv–lxvii for more information). Although UTF-8 is currently the default *charset* value of XML documents, it is reasonable to require explicit indication of the *charset* in the header of XML documents intended for import and export.

4.3 Collation

In this section, we discuss advanced data management topics related to the correct sorting of data. The discussion is aimed mostly at developers and administrators of terminology management systems and terminological collections.

Collation is the process and function of determining the sorting order of text strings. One common type of collation is alphabetization, but collation is not limited to ordering the letters of an alphabet. Collation is not uniform, being influenced by culture conventions, domain conventions, lexicographic conventions, lexicographers' preferences and technical restrictions. Collating conventions may also change over time.

In computer systems, each letter is assigned a unique numeric code, but the proper and customary ordering of strings is not performed by a simple numeric comparison of those codes. Unicode provides a specification on how to compare two Unicode strings and also supplies the *Default Unicode Collation Element Table*, in which the default collation order for all Unicode characters is specified (see Davis & Whistler 2012). This collation algorithm is implemented in all Unicode-aware database management systems.

However, the default *Unicode collation algorithm* (UCA) used in database management systems is not always sufficient for arranging lexical data in accordance with the expectations of compilers and users of reference products . Customization of collation rules is a very useful advanced feature of a terminology management system. Developers should extend the range of options for collation customization and make sure that the documentation and user help on this complicated issue are sufficient.

4.3.1 Standard-based customization of collation rules

Tools for customization of collation rules are usually table-based. The purpose of the collation table is to do the mapping from collating elements to weighting elements that can be processed by computer by the means of simple numeric comparison. The *Common Template Table* specified in ISO 14651:2007 describes four levels of comparison. However, the number of levels can be extended or reduced (ISO 14651:2007: 5).

Levels of comparison are applied in a particular order. In the ISO model, the first level determines the order of common letters of the alphabets for the given script, and the second level breaks ties on strings with and without diacritical marks. The third level resolves issues with case-sensitiveness and the fourth, optional level deals with non-alphabetical characters (ISO 14651:2007: 5, 40–41).

Some commercially available terminology management systems, such as *SDL MultiTerm* (http://www.sdl.com/products/sdl-multiterm/desktop.html, accessed 30.1. 2012) and *tlTerm* (http://tshwanedje.com/terminology, accessed 30.1.2012), provide collation customization tools. For example, in *MultiTerm*, users can specify whether they want sorting to be case-sensitive and whether non-alphabetical symbols should be taken into consideration in the sorting process.

tlTerm, in its turn, uses the full four-pass sorting based on ISO 14651:2007. Users may configure the sorting by adding or removing characters and changing their default order. The fourth level does not seem to be configurable in *tlTerm*. *tlTerm* also supports external sorting plug-ins, such as radical / stroke count for Chinese. Configurations created and customized by the users can be saved and loaded. Standard-based customization opportunities provided by *tlTerm* and *MultiTerm* are a good start, but in many cases an even more refined mechanism is needed.

4.3.2   Extensions to the collation customization table

The need for extensions to the four-level collation table is implied in ISO 14651:2007 (e.g., Annex D, section D3) and the *Unicode Technical Standard* (Davis & Whistler 2012, e.g., sections 1.4, 3.6 and 5). In this section, we describe a number of useful extensions using the experience obtained in our PhD project (Kudashev 2007), several dictionary projects (e.g., Suomalais-venäläinen metsäsanakirja 2008; Kudasheva & Kudashev 2008) and especially while developing our in-house terminology management system *MyTerMS* (Kudashev & Kudasheva 2006).

The proposed extensions are numbered. First, we make some general proposals that do not relate to any of the four levels of collation and then discuss level-specific extensions.

General proposals

1. Since the order of applying the levels of comparison is important, it should be possible to change this order, especially if additional levels of comparison are introduced.

2. It should be possible to temporarily disable individual levels of comparison or groups of levels and re-enable them when they become relevant (cf. Davis & Whistler 2012, section 1.4). Disabled levels may be shown in grey.

3. The length of strings should be used as a sorting criterion for strings that are identical characterwise. Shorter strings normally precede longer ones in lexical reference resources; for example, the headword *car* should come before *carbonate*.

4. In lexical reference resources, collation of identical strings may depend on the contents of other fields. For example, in the case of homonyms, nouns are usually placed before adjectives and verbs. Homonyms should also be sorted according to homonym indexes. Taking the contents of other relevant fields into account may require an extra sorting level or levels in the collation table.

Extensions to Level 1 (basic sorting)

5. By default, characters of other scripts are ignored during sorting at Level 1. Meanwhile, some LSP designations may consist exclusively of foreign characters (e.g., a biological Latin name for which there is no domestic term) or have them as inserts (e.g., the Russian legal terms *признание de jure* and *признание de facto*). In order to get such strings sorted in a predictable manner, an extra level or levels needs to be introduced regulating the preferred order of scripts and sorting preferences for these scripts (cf. Davis & Whistler 2012, section 1.4).

6. In some languages, two characters (digraphs) are considered as one letter and vice versa. For example, in Spanish, *ch* and *ll* were treated as single letters until 1997; in Slovak, *ch* collates after *c*; in German, *ß* = *ss*. There are also variations, such as German *ä*, which may be collated either as an accented form of *a* or as *ae*.

Since contractions and expansions of this kind are not dealt with in the *Default Unicode Collation Element Table*, special tailoring is required for them (Davis & Whistler 2012, sections 3.6). Preparatory conversions have to be made *before* the comparison of basic characters at Level 1.

Extensions to Level 2 (accents/diacritics/tildes)

At Level 2, ties are resolved on the 'secondary' letters containing diacritical marks, accents, tildes, etc. It should be noted that in some languages, letters with diacritics are considered basic letters of the alphabet (e.g., *ñ* in Spanish). Such cases should be dealt with at Level 1 rather than Level 2 (ISO 14651:2007: 5).

Quite often, there are no clear rules for collation of diacritical characters in languages where such characters can be found only in loan words (ISO 14651:2007: 40). One possible solution to this problem is that diacritical characters should follow their non-diacritical versions (e.g., the verb *resume* should come before the noun *résumè* in English). As for the precedence of different versions of diacritical

characters (e.g., é and è), the rules for their collation should be borrowed from the language from which the loan word in question originates.

7. In many languages, the first accent difference determines the final order. However, in French and several other languages, it is the *last* accent difference that determines the order, i.e., *backwards* tailoring is used to resolve ties between homographs with different diacritical marks. A priority order is assigned to each type of accent. According to this, *coté* should be sorted after *côte* but before *côté* (ISO 14651:2007: 40). This suggests that such a feature as *directionality* should be introduced into the Level 2 collation table.

Extensions to Level 3 (case-sensitiveness)

The third level breaks ties for homographs that differ only because of upper/lowercase characters. Traditions vary between languages as to whether lower- or upper-case letters should come first. For example, in German dictionaries lowercase always precedes upper case, while in French dictionaries capitals generally come first, though this is not an established rule. English does not have a monolithic practice in this respect either (ISO 14651:2007: 40–41). This implies that in many cases users may be given the freedom of choice about whether they prefer upper- or lower-case characters coming first.

Extensions to Level 4 (non-alphabetical characters)

Non-alphabetical characters are often ignored in the sorting process. However, this makes the order of the following pairs of headwords arbitrary:

(6)	coop	Adogen 142D	Aldo 75
	co-op	Adogen 141D	Aldo-37

In some LSPs, the use of non-alphabetical characters is quite extensive. The practice in how strings like 1hg, hg1, hg^{-1}, hg$_\infty$, √hg are sorted varies from one domain and dictionary to another (Landau 2001: 109). Several extensions to Level 4 are proposed below.

9. Treatment of the space character is a very important issue in LSP collections. There are two main approaches: *word-by-word* sorting when the weight of the space character exceeds the weight of alphabetical symbols, and *letter-by-letter* sorting, when space characters are ignored. *Table 5* demonstrates the difference between the two approaches.

Word-by-word sorting	Letter-by-letter sorting
ad idem	ad idem
ad litem	adjoin
adjoin	adjourn
adjourn	ad litem

Table 5. Word-by-word and letter-by-letter sorting.

Both approaches have their strengths and weaknesses, and tastes differ in this respect (see Bergenholtz & Tarp 1995: 192; Kudashev 2007: 373). The dash mark is another character that is sometimes treated in the same way as the space character. In the collation table, it should be possible to specify whether spaces and dashes should be ignored or assigned greater weight than letters.

10. Numbers are another important type of non-alphabetical symbol. By default, numbers have greater weight than letters. However, numbers are often considered less significant than letters in dictionaries. One such example is chemical compound names with prepended numerals, e.g., *1,2-diclorobenzol* (ISO 14651:2007: 37).

Numbers should not be completely ignored in the sorting process, because users expect strings that differ only in numerals to be sorted according to these numerals, e.g., *Adogen 141D*, *Adogen 142D*. In the collation table, it should be possible to specify whether numbers have more or less weight than letters (cf. ISO 12199:2000: 10–11).

11. Other non-alphabetical characters may usually be ignored during the sorting process but exceptions are still possible. Besides, it may be necessary to sort non-alphabetical characters vis-à-vis other non-alphabetical characters. This means that it should be possible to deviate from the default Unicode order with respect to non-alphabetic characters as well.

4.4 Domain classification

4.4.1 Reasons for using domain classification in a term bank

A specific characteristic of term banks is that they contain LSP designations from multiple domains. Indication of the domain to which a particular LSP designation belongs has several benefits.

Domain labels indicate the area of usage of LSP designations and give a clue about their meaning, which is particularly important in situations where proper semantic description, such as a definition or note, is not provided. In electronic collections, domain labels play a very important role in disambiguation – selection of the required headword from a list of homonyms. Domain classification also allows us to organise terminological records thematically and manage them in a systematic way.

In a collaborative terminology management system, domain classification can be an important means of managing user rights and roles. User editing rights may be restricted to only those records which correspond to the domains of expertise specified in the user profile. The ability to check whether content creators are competent in the given domain also allows users to estimate the reliability of the data better.

The results of several surveys show that users support the inclusion of a domain classification in terminological reference products. For example, translators consider domain labels as one of the most important types of information in LSP collections (Kudashev 2007: 294; Šajkevič & Ubin 1988: 115). In a survey conducted by the Swedish National Term Bank (Rikstermbanken) in 2011, over 86% of users considered adding a domain classification useful (Nilsson 2012).

4.4.2   Overview of existing domain classifications (case: Finland)

Domain classifications are used in many areas of application, for example, in statistics, planning, accounting, for classifying publications and other documents according to their principal subject, and so on. Domain classifications form part of library classifications, thesauri and upper-level ontologies.

In Finland, the most widespread library classification is YSA (*General Finnish Thesaurus*, see http://fi.wikipedia.org/wiki/YSA, accessed 16.1.2013). YSA has been recently ontologized, and its revised and extended version has become YSO (*Finnish General Upper Ontology*, see http://onki.fi/en/browser/overview/yso, accessed 16.1. 2013). Some libraries use other classifications, for example, the *Helsinki City Library* uses HKLJ (*Helsinki City Library Classification*, see http://hklj.kirjastot.fi/en-GB, accessed 16.1.2013).

Economic activities are classified by *Statistics Finland* according to the *Economic Classification* (see http://www.stat.fi/meta/luokitukset/index_talous_en.html, accessed 16.1.2013), and fields of science and technology according to the *Field of Science and Technology Classification* (see http://www.stat.fi/meta/luokitukset/tieteenala/001-2007/kuvaus_en.html, accessed 16.1.2013).

At present, many classifications are either translations or localized versions of international classifications originating from different international organizations and consortia. The source language of these classifications is usually English. An example of an international library classification is the *Universal Decimal Classification* (UDC) developed by the *UDC Consortium* (see http://www.udcc.org, accessed 16.1.2013).

The above-mentioned *Economic Classification* used in Finland is based on the Eurostat *Statistical Classification of Economic Activities in the European Community*

(NACE Rev. 2, http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/ publication?p_product_code=KS-RA-07-015, accessed 16.1.2013), which itself is based on the United Nations' *International Standard Industrial Classification of All Economic Activities* (ISIC Rev. 4, see http://unstats.un.org/unsd/ cr/registry/isic-4.asp, accessed 16.1.2013). *Field of Science and Technology Classification* is based on the *Recommendations Concerning the International Standardization of Statistics on Science and Technology* by UNESCO (see http://unesdoc.unesco.org/images/0008/ 000829/082946eb.pdf, accessed 16.1.2013).

4.4.3 Overview of domain classifications used in term banks

As the range of authoritative classifications is rather broad, the question of which ones of them is the most suitable for the purposes of terminology management arises. The fact that different term banks use different classifications although this complicates the exchange of terminological data implies that there is no simple answer to this question.

Some term banks adopt external domain classifications as such. For example, *IATE* (EU inter-institutional terminology database, see http://iate.europa.eu, accessed 16.1.2013) uses *Eurovoc* (see http://eurovoc.europa.eu, accessed 16.1.2013), which is a multilingual, multidisciplinary thesaurus covering the activities of the EU and the European Parliament in particular.

In other term banks, tailored versions of existing classifications have been adopted or dedicated in-house classifications created. For example, the Canadian *Termium*, which is one of the biggest and oldest term banks in the world, uses a sophisticated domain classification developed at the University of Montreal (Hutcheson 2001: 670).

In smaller term banks, such as the Finnish *TEPA* (see http://www.tsk.fi/ tepa/netmot.exe?UI=engr, accessed 16.1.2013) or Swedish *Rikstermbanken* (see http://www.rikstermbanken.se/rtb/mainMenu.html, accessed 16.1.2013), no domain

classification is used. As these term banks are mostly collections of glossaries, the name of the glossary usually doubles as a domain label.

4.4.4 Problems with existing domain classifications

Researchers who studied the applicability of library and documentary classifications to the needs of term banks in the 1980s (e.g., Nedobity 1988; Lingvističeskaâ koncepiâ 1989: 54), came to the conclusion that existing library classifications and thesauri might provide a good starting-point, but in most cases could not be used as such for the purposes of terminology management. A few examples of issues that complicate the use of existing classifications in terminology management systems are provided below.

1. Upper level classes of library classifications may refer to several domains, some of which are not even closely related. Such classes cannot be used as domain labels as they are too broad (and often too long as well). For example:

- HKLJ, class 630: *Metal industry. Wood processing industry. Electrotechnology. Industry textile. Leather industry*.

- Library of Congress Classification, letter G: *Geography. Anthropology. Recreation*.

2. Quite often, classes in thesauri, ontologies and library classifications are themselves *keywords* (terms) rather than *names of domains*. For example, such classes in YSO as *heart*, *ECG*, *myocardial infarction* are *terms* that belong to the domain of *cardiology*.

3. Many classes in library classifications, thesauri and top ontologies are superfluous from the point of view of terminology management. Examples include:

- Proper names (e.g., YSO contains over 200 names of computer programmes).

- Classes of publications by their genre or language in library classifications (e.g., HKJL, class 050: *General periodicals*; subclass 051.1 *Finnish-Swedish periodicals*).

- Abstract classes in ontologies (e.g., *abstract-concrete*, *endurant-perdurant* in YSO).

4. National classifications are rarely available in more than two or three languages, and they tend to be culture- and/or language-specific, at least partially. For example, YSA contains such culture-dependent keywords related to the Finnish educational system as *lukio* (≈ upper secondary school), *lyseo* (≈ secondary school) and *kansakoulu* (≈ elementary school). In Finnish classifications, one will find *valtiotiede*, which is a partial equivalent of *politology*, or *political science*.

5. Some thesauri and classifications (e.g., UDC) are too complicated for the general public, as they target professionals. Quite often such classifications are not distributed free of charge, and it is difficult to obtain usage rights to them.

6. Version management is a heavy burden of many classifications and thesauri. For example, a conversion table between two minor versions of *EuroVoc* is over 200 pages long.

In the 1980s, the need for a domain classification specifically designed for the purposes of terminology management was discussed actively (see Nedobity 1988), but such a classification seems to have never been created or at least made public. As indication of domain is a very important tool of quality assurance and role management in collaborative terminology management, it was decided to create a dedicated domain classification in the *TermFactory* project. The first step in designing the classification was to formulate the requirements and to identify probable challenges and pitfalls.

4.4.5 Requirements for domain classification

To suit the needs of multilingual collaborative terminology work, domain classification should ideally meet the following requirements. The classification:

- Should be free and available online round-the-clock, which in practice means that it should be a part of the system rather than a third-party resource.

- Should be multilingual.
- The categories in the classification should be widely acknowledged.
- May not be too culture-specific.
- Should be user-friendly and have simple organization and notation rules.
- Should be extensible, i.e., users should be able to add subclasses to it.
- Should have version management, so that older data could be made compatible with later versions of classification.

4.4.6 General problems in compilation of domain classifications

While working on the principles of domain classification for the *TermFactory* platform, we identified several general problems in the compilation of domain classifications. The first major problem is multiple alternative bases for classification resulting from the fact that domains can be classified in many different ways. For example, *astronomy* can be classified according to the physical bodies that are the objects of observation (e.g., solar / stellar / galactic astronomy) or according to the observed region of the electromagnetic spectrum (e.g., radio / infrared / optical astronomy). *Construction* can be classified according to the type of object to be constructed (e.g., buildings, bridges, roads), according to the stages of construction work (from design to finishing) or according to the specialization required (e.g., demolition, installation). The choice between various bases of division is not always easy (cf. Hutchenson 2001: 671).

The second challenge is the choice of appropriate depth of the classification. Granulation of domain classifications varies from one to up to nine levels of hierarchical relations (ISO 12620:1999: 23). Shallow classifications are uninformative and sometimes even misleading (cf. Bergenholtz & Tarp 1995: 153), whereas overly detailed classifications are hard to use and maintain (cf. Ubin 1992: 55; Grinev 1995: 85). Detailed classifications also tend to be subjective.

The third problem is related to the life-cycle of the disciplines. At the end of the 20th century, the number of scientific disciplines used to double every 25 years (Grinev 1993: 8), and the pace has only increased since then. This poses several problems, such as how to guarantee that the classification is comprehensive at the time of its compilation, how to keep the classification up-to-date in the future and how to know whether a new discipline is going to become well-established or is just a buzzword.

The fourth major problem has to do with the fact that classification schemes created in different countries and different languages may differ both in terms of the content of classes and in terms of their location in the classification scheme. For example, the relations between Russian *машиностроение* (a loan translation and a close relative of the German *Maschinenbau*), Finnish *metalliteollisuus* and English *mechanical engineering* are quite complex, although the final product of these industries is often the same.

Yet another problem is synonymy. Names of domains may have variants and synonyms, such as *animal geography – zoogeography*; *legal history – history of law*. If synonymous names are not included in the domain classification, users may not be able to recognise the domain under a different name. If synonyms are included but not clustered properly, LSP designations belonging to the same domain will end up in formally different domains.

4.4.7 Principles of compilation of the TermFactory core domain classification

TermFactory core domain classification was created with the above-mentioned challenges and requirements in mind. While it is based on several existing classifications, thesauri, ontologies and encyclopedias (please refer to *Appendix 3.5* for the list of primary reference sources), it is more than a compilation from several

sources. In this section, we describe the main principles of compilation of the classification.

The use of the core domain classification is supposed to be mandatory in the *TermFactory* platform, i.e., users have to link each LSP designation to at least one domain class. To speed up the process of adding domain labels, the system may allow storage of 'favourite' domain labels in user profiles as well as specifying the 'preferred domain' that will be readily selected in each newly-created term record (cf. Kudashev & Kudasheva 2006).

Number of classes

TermFactory core domain classification contains about 700 classes and is intended to cover all domains of knowledge and activity. The benefit of having a relatively small domain classification of top-level classes is that these classes are more stable and less culture- and language-dependent than classes at deeper levels. Besides, a compact classification is easy to browse and navigate.

Hierachy levels and their relations

The paper version of the core domain classification looks like a two-level hierarchy. The top level consists of about 100 classes, most of which are subdivided further. Below is an example of the *Physics* domain:

(7)   **Physics**
acoustics; atomic physics; biophysics; geophysics; particle physics; quantum physics; mechanics; molecular physics; optics; electrodynamics; thermodynamics; nuclear physics.

However, TermFactory domain classification was conceived from the very beginning as an ontological resource, and as such, it has certain pecularities as opposed to a strict hierarchy. In particular, TermFactory core domain classification makes use of the 'open world assumption' characteristic of many ontology languages (see http://en.wikipedia.org/wiki/Open_world_assumption, accessed 17.1.2013). To our mind, this makes the domain classification more flexible and user-friendly.

Each node in the classification has its own URI, i.e., constitutes a complete classifier alone. However, classes at the second level of hierarchy are also considered subclasses of the corresponding upper classes. This allows intelligent searches at the upper level of the classification as well. For example, if a user has labelled an LSP designation as belonging to *Acoustics*, which is a subclass of *Physics*, then a search for designations related to *Physics* will by default return the results related to *Acoustics* as well. As an advanced search option, users may choose to exclude subclasses from the search.

If users are not sure whether they should link an LSP designation to a more specific or more general class, they should be prompted to choose the more general one. Specifying a broader domain is less problematic than narrowing it too much.

Second-level classes may appear in the classification under several top-level classes. For example, *Legal history* can be found under both *History* and *Law and legislature*. This helps users to locate domains of an interdisciplinary character in the domain selector more quickly.

The fact that some classes can be found in several places in the domain classification does not affect the way in which they are documented in the *TermFactory* platform. In the example above, the domain label will always be *Legal history* and not *History: Legal history* or *Law and legislature: Legal history*.

A top-level class can also be a subclass of another top-level class in the TermFactory domain classification. For example, *Zoology* is a subclass of *Biology* but, being a big domain, it is also a top class in its own right with a number of subclasses.

In a strict hierarchical classification, such tricks with domain classes would be illegal. However, in a logic based on the 'open world assumption', they are perfectly appropriate as long as the domain classes are not explicitly declared disjoint from each other.

Treatment of disciplines of a broad nature

The nature of some disciplines is so broad that they can be combined with almost any other domain. For example, such words as 'philosophy', 'history', 'politics', 'sociology' or 'psychology' can be added to almost anything. However, only the major and most important branches of the corresponding sciences could be included in the core domain classification, cf. the *Philosophy* domain:

(8) **Philosophy**
axiology; aesthetics; ethics; philosophical trends; philosophy of history; philosophy of language; logic; metaphysics; philosophy of mind; legal philosophy; ontology; political philosophy; philosophy of science; epistemology; philosophy of religion; social philosophy.

Multiple domain labels

Users can label an LSP designation as belonging to several domains in the TermFactory domain classification. For example, the term *nature tourism* can be classified as belonging to *Forms of tourism* and *Leisure and hobbies*; *fuel wood* may be labelled as belonging to both *Energy production* and *Logging*. It should be noted, however, that if the same object is considered in different disciplines from different points of view (for example, if *fuel wood* is defined as a *source of energy* in the energy sector and *timber assortment* in the forestry sector), it is advisable to consider such LSP designation as homonymous (related to two different concepts) and provide them with different domain labels.

Treatment of complex disciplines

Some long-established areas of research and activities are complex. For example, *Marine research* is a complex conglomerate of individual disciplines, such as hydrology, geology, geography, meteorology, marine biology, and so on. Such complex and somewhat loose conglomerates are split into individual disciplines in the TermFactory domain classification.

Treatment of culture- and language-specific domains

Since the TermFactory core domain classification aims to be as generic and international as possible, the culture-specific divisions and classes were avoided. The general practice upon encountering culture-specific classes was to find the lowest common denominator. For example, the problem mentioned above with the Finnish *valtiotiede*, which is a partial equivalent of *politology* (or *political science*), was solved by using a more general class – *political research*. When a choice between two or more culture-specific divisions was inevitable (e.g., what system to follow in dividing *Law and legislature* – civil law, common law, religious law, etc.), priority was given to the divisions adopted in continental Europe.

Complex classes

In some library classifications, a complex domain class may include up to a dozen individual classes. Sometimes the classes are also rather loosely related. For example, in the *Universal Decimal Classification*, the last component of the class *Public health engineering. Water. Sanitation. Illuminating engineering* is quite distant from the first three.

In the TermFactory domain classification, combining of classes is allowed only when they are closely related to each other. Examples include *Administration and management*, *Ethnology and ethnography*, *Cosmetology and beauty services*. The maximum number of domains making up a single class is restricted to three.

Treatment of interdisciplinary designations

Interdisciplinary words and word combinations like *analysis*, *report*, *evaluation*, *document* or *method* can be placed in the TermFactory classification into the *General terms* class. One should keep in mind that such designations might also have a specific meaning in various LSPs. For example, in programming, the word *document* has a specific meaning that differs from the general interdisciplinary meaning.

LSP designations of general character may also coincide with the short forms of terms. For example, *analysis* may appear in chemical texts as a short form of *chemical analysis* and in mathematical texts as a short form of *mathematical analysis*. These short forms should be labelled as relating to *chemistry* and *mathematics* correspondingly.

Treatment of variants and synonyms

Common variants, synonyms and near-synonyms of the primary names of domain classes are provided in the TermFactory domain classification in brackets; for example:

(9)   Industrial and organizational psychology (<- I-O psychology; industrial-organizational psychology; work psychology; organizational psychology; work and organizational psychology; industrial psychology; occupational psychology; personnel psychology).

In the domain selector, cross-references may be used to link variants and synonyms to the primary form. Inclusion of synonyms allows users to locate the required domain in the domain selector more easily.

Support of multilingualism

In the *TermFactory* project, the core domain classification was compiled in four languages: Finnish, English, Russian and German (please see *Appendices 3.1–3.4*). Other language versions are supposed to be produced collaboratively in subsequent projects. Finnish was the source language of the domain classification, so it is advisable to use the Finnish version as a basis for translations into other languages.

If the core domain classification is not yet available in a particular language, it is advisable to use the English language version temporarily instead. Labels in English can be later converted automatically into the required language when the core domain classification becomes available in that language.

As each domain class has a unique identifier (URI), mapping between the language versions can be done automatically. By default, domain labels are displayed in the same language as the LSP designations to which they relate. However, users may be allowed to select the preferred language in which they want the domain labels to be displayed. These preferences may be stored in user profiles.

Syntactical and typographical conventions

In library classifications and especially in thesauri, the names of classes are often given in the plural (e.g., *devices*, *castles*, *ports*, etc.). In the TermFactory domain classification, the names of the classes are usually given in the singular, except for *pluralia tantum* and collective classes, such as *Studies of science*, *Forms of tourism* or *Technical services*.

In the paper version of the TermFactory domain classification provided in *Appendices 3.1–3.4*, the following conventions were adopted:

- The names of upper-level classes are bolded and start with a capital letter so that they are visually separated from lower-level classes.
- The names of lower-level classes are not highlighted and start with a small letter (unless grammatical rules require otherwise).
- Classes are arranged alphabetically in the Finnish version and enumerated for the purposes of alignment with other language versions.
- Synonyms are provided only in connection with the main name of the class but not as cross-references so that they do not interfere with the enumeration.

4.4.8 User extensions to the core domain classification

Users may supplement the TermFactory core domain classification with their extensions following the principles described in this chapter. Domain labels from the core domain classification and user extensions should be stored in the system separately. Users may pick extensions already created by other users or add extensions of their own.

Users should attach their extensions to the second level of the core domain classification rather than the top level whenever possible. For example, if the user wishes to add the extension *Fluid mechanics*, he or she should add it to *Mechanics*, which is a subdivision of *Physics*, rather than directly to *Physics*. The language of the extensions must correspond to the language of the LSP expression being described.

Before making extensions, users should make sure that the 'missing' subclasses are not traditionally classified under some other class. Many classifications, for example, NACE and its localizations, provide detailed descriptions of domain classes with search possibilities. Such descriptions help find the right place for domains that are not explicitly present in the classification.

Extensions should be based on major existing classifications, thesauri and ontologies in the first place. The list of the primary reference sources of the TermFactory core domain classification (see *Appendix 3.5*) serve as a good starting-point. The source of extensions, their relation to the original classes (e.g., truncation, abbreviation and other modifications) as well as intended coverage should be carefully documented whenever possible. Providing the same information for the core domain classification is a valid topic for a future collaborative project.

If users feel that none of existing classes is suitable for their extensions, they may post a question to the *General Help on the TermFactory Platform* forum. If their assumption proves true, the TermFactory administrator may establish a new class under a special *Unclassified domains* class or one of its subclasses – *Unclassified field of special knowledge* or *Unclassified activity field*. Classes placed there will be taken into account in the next revision of the core domain classification.

4.5 Data category classification

In this section, we deal with topics related to data category organization in terminological collections. We discuss the concept of *data category*, provide typical examples and a classification of mismatches between data categories in various collections, and propose a linguistic classification of data categories that may serve as a bridge between terminological and lexicographical collections with different structures. Among the primary applications of such a classification are the organization of advanced full-entry search and customization of entry views in multiple collections.

4.5.1   The concept and function of data categories

In a reference product, data is split into blocks called data fields. A *data category* is the result of specification of a given data field (ISO 1087-2:2000: 13). For example, data category the 'part of speech' data category is a generalized image of all 'part of speech' data fields, including such specifications as definition, representation, permitted values, and so on.

The minimal set of data categories in a reference product includes the object of description ('headword') and the description related to it. However, in practice, the description is usually split further into smaller data categories.

A popular metaphor for explaining the concept of a data category is that of a wardrobe in which different drawers are used to store different types of clothes. This metaphor also helps us understand the function of data categories. One of the main reasons for using a wardrobe is to organize clothes in a neat way for quick and easy retrieval. Data categories play almost the same role, allowing users to retrieve and manage similar types of data conveniently.

Wardrobes come in different styles and sizes, as users have different needs and preferences as regards storing and sorting their clothes. Similarly, data category sets

and classifications vary from one terminological collection to another, and their global unification is hardly possible or desirable. However, the use of different data category sets complicates data retrieval from multiple collections and diminishes interoperability between them.

Lack of interoperability is the price that has to be paid for flexibility and freedom of choice in each terminological project and, in many cases, the problem cannot be fully solved. However, interoperability between heterogeneous collections can be improved with the help of mapping. This can be done directly between two or more collections, but a more effective and sustainable way is to do the mapping via an intermediate data category inventory or classification. This removes the need to do the mapping again for each new combination of collections. In order to be able to choose the right type of mapping for different occasions, we first need to look at the core of the problem and examine the most common types of mismatch between data category sets.

4.5.2   Typical mismatches between data categories

Data categories are the result of data classification. Data may be classified in many different ways depending on the views of the classifier and the needs of the end users. Below are a few examples of typical mismatches between data categories in terminological databases:

- Mismatch of the names of data categories due to synonymy and different naming rules.

(10)   *Example*: the same data category may be called *note* in one collection, *comment* in another one, and *NB* in a third.

- Mismatch of the contents of data categories due to a different language or notation.

(11) *Example*: the part of speech 'noun' may be coded in various collections as *noun*, *n*., *subst*., etc.

- Mismatch of the contents of data categories due to homonymy or different interpretations of their names.

(12) *Example of intralingual homonymy*: the data category *synonym* may be defined (or interpreted) as corresponding to 'full synonym', 'near-synonym' or 'full or near-synonym'.

(13) *Example of interlingual homonymy (false friends)*: the English *abbreviation* and *acronym* (as defined in ISO 12620:1999: 6–7) and Russian *аббревиатура* and *акроним* (as defined in Grinev 1993: 153) correspond to each other crosswise.

- Mismatch of the 'sizes' of data categories due to different granulation.

(14) *Example*: the data category 'abbreviated form of term' is split into five subclasses (*abbreviation*, *short form of term*, *initialism*, *acronym* and *clipped term*) in ISO 12620:1999 but there is no such division in a 'lighter' standard ISO 12616:2002 ('Translation-Oriented Terminography').

- Overlapping of data categories.

(15) *Example*: the *example* and *context* data categories overlap. Some examples are contexts, and some contexts may serve as examples, but these categories are not identical.

- Mismatch of the 'places' of data categories, i.e., their location in the classification scheme.

(16) *Example*: the *context* data category is considered concept-related data in ISO 12620:1999, apparently because contexts may provide additional information on the concept. However, since a more common function of contexts is to provide information about term usage and collocations, 'some databases classify *context* as a term-related data category' (ISO 12620:1999: 25).

- Mixed cases.

(17) *Example of a cross-language overlapping and mismatch of the 'sizes':* in ISO 12620:1999, labels for describing LSP expressions belonging to the 'lower style' include *slang register* and *vulgar* register, which partially correspond to the concepts *профессиональный жаргонизм* (professional slang) and *профессиональное просторечие* (professional colloquialism) in Russian (see Šelov & Leičik 2012: 53–69). In Finnish, however, there is only one category, *ammattislangi* (professional slang, see Sanastotyön käsikirja 1989: 12).

Mismatches between data categories may be divided into three main groups. In the first case, the definition of data categories is more or less the same, and differences only concern the name of the category or the presentation style of its values. In the second case, the data categories themselves are different (larger, smaller or overlapping). The third group includes mixed cases when both data categories and their means of representation are different.

4.5.3 Mapping between formally different data categories

When differences between data categories are only formal (different naming rules, language, or notation), they can be mapped with the help of a data category reference inventory. For example, the 'part of speech' data category and the 'n.' value permitted in it can be mapped onto a 'POS' data category and a 'subst.' permitted value through a 'partOfSpeech' reference data category and the permitted value 'noun'. Such mapping can also be done across languages.

One of the reference inventories that can be used for the mapping of formally different data categories is the ISOcat data category registry (http://www.isocat.org). The ISOcat initiative provides a standard-based framework for collaborative defining of data categories in the domain of linguistics. Its terminology section comprises several hundred data categories primarily borrowed from ISO 12620:1999.

4.5.4 Mapping between structurally different data categories

When differences between data categories are structural, i.e., they differ in terms of their definition, mapping has to be done via a common denominator. When data categories differ in size, and smaller data categories are fully covered by larger ones, the largest data category becomes the common denominator. For example, if in one collection the 'abbreviated form of term' data category is not split further but is split into several data categories in another collection ('abbreviation', 'initialism', etc.), the larger category will serve as the common denominator. However, in the case of overlapping and in mixed cases, a superordinate category is needed to cover both or all of the overlapping categories.

In order for the system to find a common denominator for two or more data categories from different collections, a hierarchical bridging classification of data categories is required, and mapping has to be done from each collection to this classification.

A bridging classification of data categories may not be too deep, as its categories are supposed to be larger than 'primitive' data categories that are not subdivided further. In ISO 12620:1999, the deepest level of hierarchy for term- and concept-related data categories is level 4 (e.g., term-related data categories -> term type -> abbreviated form of term -> initialism). In 'lighter' standards (such as ISO 12616:2002 – 'Translation-Oriented Terminography') as well as in most real-life terminological collections, the deepest level of hierarchy is level 3, which means that the bridging classification should not have more than two and a maximum of three levels. This degree of precision is sufficient for the purposes of information retrieval and customization of entry views in the vast majority of cases.

An important requirement for a data category bridging classification is that it should be intuitive and user-friendly, so that users of term banks are able to quickly and correctly specify what kind of data they are interested in when they perform a

search or customize entry views. To our mind, the ISO 12620:1999 data category classification and its modifications do not fully meet this requirement. We discuss this question in more detail in the next section.

4.5.5   ISO 12620 data category classification

The 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 (ISO 12620:1999: 1). The standard contains a hierarchical list of data categories together with their definitions. ISO 12620:1999 was superseded by ISO 12620:2009, but the latter standard no longer provides a fixed list of data categories, specifying instead the principles of organization and management of a collaboratively created Data Category Registry (DCR). The practical realization of ISO 12620:2009 is the ISOcat DCR (see http://www.isocat.org).

At the same time, apart from some omissions, additions and rearranging, the terminology section of the ISOcat DCR is still largely based on the set of data categories from ISO 12620:1999, a classification also widely used in its original form. For example, the default data category set of the TBX standard (ISO 30042:2008 – 'TermBase eXchange') is primarily based on ISO 12620:1999 (ISO 30042:2008: 2).

Meanwhile, the ISO 12620:1999 data category classification is problematic in many respects. To start with, there are some inconsistencies concerning the principal division of data. According to section 6.2 (*Typology of data categories*), data categories are divided into three main groups in ISO 12620:1999: term and term-related information, descriptive data and administrative data. However, in Annex D *(Systematic listing of data categories)*, the second group is called *Data categories related to concept description*. At the same time, this group also contains the *Note* subgroup, which is supposed to stand alone 'because it can be associated with any

104

one of the other categories and therefore cannot be subordinated to any other specific subgroup' (ISO 12620:1999: 4).

Even if we assume that the intended division included four groups: term and term-related information, concept-related information, administrative data and *Note*, the classification still raises many questions. To mention a few:

- Why are *examples* and *contexts* concept-related data and not term-related data? Cf. the description of the *context* data category: 'A text or part of a text in which a term occurs' (ISO 12620: 1999: 25).

- Why are *synonym* and *equivalence* term-related data while everything otherwise related to the meaning is concept-related data? Cf. description of the *degree of equivalence* data category: 'The extent to which the intensions of two or more concepts overlap' (ISO 12620: 1999: 21).

- Why is *antonym* administrative data and not term- or concept-related data?

- Why are *non-textual illus*trations (*audio*, *video*, etc.) merely concept-related data?

- What is the exact definition of *administrative data*? Why does this class include such heterogeneous categories?

The division of terminological data into concept-related and term-related may be useful from the technical point of view because it supports the concept-oriented approach that reduces the number of relations between terms by linking synonymous terms to the same concept. However, such a division in general and its implementation in ISO 12620:1999 in particular may present a challenge for everyday users of terminological databases – translators, technical writers and domain experts. Indeed, it is not easy to comprehend why *synonyms* should be searched for in term-related data but *antonyms* in administrative data and *examples* in concept-related data.

In the ISOcat DCR, the problem was 'solved' by removing the top-level categories, 'subgroups' (such as A.2 terms-related information, A.3. equivalence, A.4 subject field, A.5 concept-related description, etc.) but this sounds like throwing the

baby out with the bath water. Categories that were reasonably grouped thematically in ISO 12620:1999 (such as equivalence -> degree of equivalence, false friend, directionality, reliability code, and transfer comment) are now scattered around in the alphabetical order. The classification has become flat, with most categories being top-level classes. This is the reverse of what is required of a classification aimed at bridging structural mismatches.

In our opinion, the solution is not in giving up the classification but in choosing new bases for it that are more adequate. Since everyday users work with terms – words and word combinations – it is natural for them to speak about the meaning of a term, its synonyms, examples of its usage, and so forth. This implies that a user-friendly bridging classification of data categories should be term-oriented (meaning 'oriented towards one form-meaning pair', not 'form-oriented') and based on major linguistic categories, such as form, meaning, and usage. An additional benefit of a linguistically oriented classification is that it may serve as a bridge not only between various terminological collections but also between terminological and lexicographical reference resources. In the next section, we present our vision of a linguistic classification of data categories designed for the purposes of terminology management.

4.5.6   Linguistic classification of data categories

Since LSP designations are linguistic signs, information about them may be divided into information about their form, meaning, usage, relations with other designations, origin and development (cf. Kudashev 2007: 197). Besides, terminological reference products may contain elements of encyclopedic description. The top level of the hierarchy is supposed to be relatively stable, while the second level and especially the third are open sets of subclasses that can be supplemented if needed. Below we discuss each of the top-level classes and their subclasses in more detail. A compact version of the classification is provided in *Appendix 1.1*.

Data related to form

Data related to form may be divided into the following subclasses:
- data related to type of designation by its form
- data related to spelling
- data related to pronunciation
- data related to word formation
- data related to inflection.

Data related to type of designation by its form

This data category is intended for classifying designations by their form, e.g., full form, abbreviated form, transliterated form, and so on. Data in this category may relate to either written or oral form, or both. For example, indication that a particular expression is an initialism, i.e., an abbreviated form made of the initial letters of the full term, in which these letters are pronounced individually (as in United Nations – UN), provides information about both the written and oral form of the designation.

Data related to spelling

Examples of data related to spelling include the following categories:
- graphical representation of the form
- spelling variants
- hyphenation (indication of the conventional points at which words may be divided when they occur at the ends of line – Hartmann & James 1998: 70).

Data related to pronunciation

Data related to pronunciation may include such classes as
- representation of pronunciation (e.g., transcription)
- pronunciation variants
- syllabification (written representation of division of words into phonic syllables – Hartmann & James 1998: 134).

Data related to word formation

This data category may include such subcategories as
- term components and term elements
- morpheme structure
- syntactical model
- word formation model
- word formation method(s)
- morphological variants
- derivatives.

Data related to inflection

Examples of data related to inflection include:
- grammatical parameters (number, gender, animacy, etc.)
- inflected forms (for canonical form)
- canonical form (for inflected forms)
- inflection models.

Data related to meaning

Data related to meaning may be split into four subcategories:

- data related to logical meaning
- data related to type of designation by its logical meaning
- data related to induced meaning
- data related to degree of transparency of designation.

Data related to logical meaning

Logical meaning is the intention of the logical concept denoted by an LSP designation, roughly corresponding to *denotative meaning* in general linguistics. In terminological collections, data related to logical meaning is usually provided in the form of terminological definitions and definition-like descriptions.

Data related to type of designation by its logical meaning

This category is intended for classifying designations by the type of concept they denote, e.g., into terms, appellations, nomenclature, prototerms, etc.

Data related to induced meaning

By *induced meaning,* we mean various components of meaning that are not directly related to the logical concept denoted by an LSP designation. Induced meaning is a concept similar to but broader than that of *connotative meaning* used in general linguistics. Examples of components of induced meaning include:

- various connotations, i.e., evaluative components of meaning
- the inner form of designation (its 'literal', morpheme-by-morpheme meaning)
- various kinds of associations, e.g.,
    - associations resulting from consonance and rhymes
    - associations resulting from other LSP or LGP meanings of the same designation
    - associations resulting from antonymous, synonymous, paronymous and other lexical relations of the designation.

As one can see from the list above, additional components of meaning usually result from the attitude of language users towards the objects denoted by LSP designations and from user associations resulting from the relations of an LSP designation with other designations or language components. In this sense, additional components of meaning are 'induced' by language users, other designations, or both.

Induced components of meaning are welcome and even cultivated on purpose when they create positive associations and connotations or allow users to express their attitude to the subject in informal communication. At the same time, they may distract the attention of users from the logical meaning that is supposed to be at the core of LSP communication. This is probably one of the reasons why components of induced meaning have been largely neglected in traditional terminology theory.

However, these components of meaning often play a crucial role in the selection or creation of adequate designations for new concepts and in prioritizing synonymous designations. Their importance is also apparent in sensitive domains, such as politics, public relations, or social work (cf. Kudashev & Kudasheva 2010). For a discussion on the presentation of components of induced meaning in terminological reference resources, please see (Kudashev 2006; Kudashev 2007: 254–258).

Data related to degree of transparency of designation

In the ideal case, LSP designations should be 'transparent', so that their meaning is fully deducible from the meaning of their components or otherwise motivated (see Šelov 2003: 152–174 for a discussion on how this 'transparency' and 'deducibility' can be measured). However, for a number of reasons, the inner form of LSP designations may reflect its intended meaning only partly or even be misleading. Correspondingly, LSP designations may be divided into a number of classes depending on how much their inner form reflects their meaning, or how much they are 'motivated' (see Leičik 2006: 39; Kiâk 1989: 39–40).

Data related to encyclopedic description

Encyclopedic description provides information about objects denoted by LSP designations, this information being extralinguistic in most cases. However, it is sometimes not easy to draw the line between description of a concept and description of an object denoted by an LSP designation. Many terminological databases contain encyclopedic descriptions, and the proportion of such information is expected to grow in the future as the boundary lines between different types of reference product are rapidly disappearing under the impact of computer technology (cf. Hartmann 2001: 5). This justifies the inclusion of encyclopedic description in the classification.

Data related to usage

Data related to usage may be subdivided into two large categories:

- data related to restrictions in usage
- data related to frequency of use.

Data related to restrictions in usage

Usage of any LSP designation is restricted to at least some language and domain. Besides, usage may be restricted to some geographical area, chronological period, professional group, register, and so on. Below is a list of common usage restrictions:
- language restriction
- domain restriction
- geographical restriction
- chronological restriction
- scientific school/theory restriction
- organizational restriction
- proprietary restriction
- register restriction
- professional group restriction
- compliance restriction
- combinatory restriction.

The two latter types of restriction probably require a more detailed explanation. Compliance restrictions relate to the fact that terminological reference products tend to be relatively prescriptive in nature. Their normativeness may range from recommendations by the compilers or domain experts whom they consulted to normative authorization. In any case, the question is about the compliance of an LSP designation with a 'good' or 'correct' style (from the compilers' point of view), and in case of terminological standards with some normative document as well.

Combinatory power is the ability of linguistic units to be used together. Combinatory power may be divided into semantic, lexical and syntactical. Semantic combinatory power means that expressions do not have controversial components in

their meaning. Lexical combinatory power manifests itself in the ability of expressions to combine with certain other lexical means. Syntactical combinatory power is the ability of an expression to combine with certain grammatical forms of other expressions and auxiliary words. Combinatory restrictions are usually described with the help of syntactical models, examples and contexts.

Data related to frequency of use

Information about the frequency of use may be based on corpus evidence and expressed numerically, or may be a subjective estimate expressed verbally (e.g., commonly used – infrequently used – rarely used).

Data related to lexical and logical relations

This data category includes lexical relations (synonymous, antonymous, etc.), hierarchical logical relations (generic and partitive), non-hierarchical ontological relations (in a broad sense, as relations conveying knowledge about the world, such as sequential, temporal or causal relations), and cross-language equivalence relations. Relations pertaining to word formation, inflection and the combinatory power of expressions do not belong in this category.

Unfortunately, we could not find a compact name for this data category that would cover all four types of relation. Theoretically, equivalence may be viewed as cross-linguistic synonymy, but for practical reasons it is better to reserve a separate class for it. Philosophically, logical relations may be viewed as a species of ontological relation, but working terminologists tend to talk about logical relations even when they refer to non-hierarchical relations. Besides, the word combination 'ontological relations' has become ambiguous because of two different meanings in the domain of philosophy and computer-assisted knowledge modelling. In our opinion, 'lexical

and logical relations' is a working compromise that is unlikely to cause misinterpretation.

Data related to origin and development

This data category may include the following types of data:
- information on the forms from which the designation is believed to originate
- information on the stages of development of the designation
- information on earlier form(s), meaning(s), usage, etc. of the designation.

4.5.7 Mapping of other data category classifications onto linguistic classification

The primary function of the linguistic classification of data categories is to serve as a bridge between heterogeneous terminological collections. Using such a classification, users do not need to consider different names, sizes and places of data categories in different collections, but can simply specify that they are interested, for example, in information related to meaning, logical relations or usage restrictions. They can perform a search in data categories that correspond to these criteria and/or customize the entry view in such a way that only this kind of information is visible. For example, they may ask the system to display information about logical relations in which a particular term is involved or a description of its meaning in different collections. Linguistic classification can also serve as a 'subject directory' when users work with a collection the data category set and organization of which are foreign to them.

Mapping onto a bridging hierarchical classification of data categories helps overcome structural mismatches between data category classifications while mapping onto an authoritative inventory of data categories (like the ISOcat DCR)

may help fix formal mismatches. Both types of mapping may be seen as useful extensions to the basic specification of data categories and data category schemata. Enhanced data retrieval possibilities help in finding inconsistencies and redundancies and may have a considerable positive impact on the quality of terminology work.

During the mapping, a category from one classification may be mapped onto two or more categories from another classification. For instance, *example* is a typical multifunctional field that may contain information about the form, meaning and usage of an LSP designation in different proportions. For the purposes of specifying the threshold while searching, it is advisable to specify the degree of correspondence numerically (e.g., 0–100%).

If the definition of a data category is not precise enough, or multiple interpretations of the name or definition of the category exist or may be expected, mapping may involve uncertainty factors. For example, the *neologism* value from ISO 12620:1999 sounds like a chronological label while in fact it refers to the methodology employed in creating the term. Since some users may misinterpret the purpose of the data category, one may expect to find both data related to term provenance and data related to the chronological status of the term in this data category. Such uncertain cases should be properly marked so that users may choose to filter them out in the search options.

In *Appendix 1.2*, we provide a sample mapping of ISO 12620:1999 data categories onto the linguistic classification. The mapping is done only for categories that are related to terminological description. Administrative and bibliographical data categories cannot be mapped onto the linguistic classification of data categories. Their classifications are discussed in the next chapter.

5 Descriptive metadata and its role in quality assurance

Descriptive metadata helps users interpret, evaluate and improve terminological descriptions as well as navigate within terminology management entities and administer them. In this manual, we focus on two types of descriptive metadata: data related to the documentation of sources, and administrative data.

5.1 Documentation of sources

Indicating sources provides valuable information for both users and compilers of terminological reference resources. Doing so allows users to estimate the quality of terminological description. In cases of doubt, they can check the source of information and consult it directly for details. Indication of sources for compilers is useful for several reasons:

- compilers show respect for the intellectual rights of other authors and at the same time shift a large share of responsibility onto them
- sources are a strong argument in disputes with other members of the team or community
- indication of sources helps to find inconsistencies between various sources and resolve them.

The importance of documenting sources in terminology work is confirmed by the existence of a dedicated ISO standard (ISO 12615:2004 – 'Bibliographic References and Source Identifiers for Terminology Work'). There are also various formats for exchanging bibliographic data between terminological collections (e.g., TeDIF – Betz & Schmitz 1999; IETR 2005; TBX – ISO 30042:2008: 27–28).

Because of the variety of applications, various methods and levels of detail are used to record bibliographic information on terminology sources (ISO 12615:2004: 4). The more data is provided about the sources and the better it is classified, the richer the functionality of the system. Detailed description and fine granulation enable

advanced queries about various characteristics of the sources and the automatic generation of bibliography in different styles. Dedicated applications, such as source management systems, provide an example of such facilities.

However, a terminology management system, especially a collaborative one, is unlikely to reach the level of advanced source management systems. For the designers' part, providing and keeping up-to-date facilities for professional-level source management sounds like the wrong priority and an unreasonable investment. On the users' side, it is hard to expect community members to have enough energy, time and expertise to document sources with professional depth and precision.

In our opinion, the best strategy for designers of a collaborative terminology management system would be to make sure that at least a minimum set of data is provided about every source and at the same time give free rein to those users who want to go beyond the minimal requirements and follow the best practices in source documentation. In this section, we discuss both basic requirements for the documentation of sources and possible extensions to them. We also provide templates for documenting various types of sources in *Appendix 4*.

5.1.1   Types of data related to the documentation of sources

Data related to the documentation of sources in terminological databases can be divided into two big classes: *source references* and *bibliographic records* to which these references are made. In theory, full bibliographic records may be used as source references. In practice, however, bibliographic records are usually stored separately and shorter aliases of them are used in the entries. There are several reasons for doing so.

First, as bibliographic descriptions are usually quite long, shorter references to them help save space and improve the readability of the entries. Second, it is not reasonable to duplicate the full description of a source as many times as it is referred to. Third, it is easy to modify a full bibliographic record that is stored separately.

Yet another reason for keeping references and bibliographic descriptions separate is modularity. If full bibliographic records are stored separately, it is possible to provide them with annotations and split them into smaller data categories for advanced querying. For example, if the publication date of the sources is provided as a separate data category, searching for outdated sources and entries that contain references to them becomes an easy task.

5.1.2 Basic requirements for source references

A source reference is supposed to be relatively short, user-friendly and supply some information about the source. Unfortunately, short and meaningful references are unlikely to be persistent, as there may be situations in which they have to be changed; for example, because they sound funny or inappropriate in some language that was not thought of at the time of creation of the source reference. If modified, a source reference has to be updated throughout the term bank and its archived versions, which may be difficult to implement.

The problem can be solved by distinguishing between a *visible source identifier* displayed in the entries and a persistent *source ID* used by the system for an unambiguous identification of the sources and for linking them to the corresponding bibliographic records.

The source ID must be both unique and persistent, at least within the scope of the term bank. An even better choice is to use global identifiers, such as URI or ISBN. Visible source identifiers must be unique but may be modified if necessary. They should not be 'hard-coded' in the term records but generated on the fly based on the persistent source ID. This solves the problem of content updates, synchronization and versioning.

The data schema should allow the addition of one or more source references to any data field related to the terminological description. In extensive fields such as

definition, *note* or *example*, it may also be necessary to add source references within the field.

5.1.3 Additional elements of source references

In addition to the *source identifier*, a source reference may contain a number of optional fields. The most common include indication of the source type, indication of the location of the cited abstract in the original document and indication of the source usage, such as *cf.*, *as cited in* and *see also*.

Indication of the source type

A source reference may contain an indication of the source type. For example, in the *Dictionary of Lexicography* (Hartmann & James 1998), printed sources are preceded by an open book pictogram (📖) and Internet sources by a PC pictogram (💻). Source type indications may be graphic images or 'letters' of symbolic fonts, as in the example above. The use of symbolic fonts has many advantages but it may cause cross-platform incompatibility and incorrect presentation of data if the font is missing from the client computer.

Indication of the citation's location

A source reference may also be supplemented with information about the location of the cited passage in the original document. Page numbers are the most common type of location indicator. They may also be supplemented with the number of a column, paragraph, row, and so on. In highly structured documents such as laws, and in documents with no pagination, references are usually made to the logical

divisions of the text (chapters, sections, paragraphs, and so on). In some types of electronic document, references may also be made to special bookmarks or anchors.

Indication of the source usage

If the relations between the text and the source are more complex than direct borrowing, this may be indicated by special expressions or symbols. For example, indirect borrowing may be indicated by the ≈ sign placed before or after the source label (cf. this example from the *Finnish-Russian Forestry Dictionary* – Suomalais-venäläinen metsäsanakirja 2008):

> (20) **aapasuo**
> suoyhdistymä, jonka keskiosa on reunoja alempana 📖 ≈ SESMS

See references (as in *see Laine & Vasander 1990*) may be expressed with the help of an arrow pointing right: ⇨ 📖 Laine & Vasander 1990. *Cf.* references (as in *cf. Laine & Vasander 1990*) may be indicated with an arrow pointing in both directions: ⇔ 📖 Laine & Vasander 1990. The *as cited in* type of references (e.g., *Laine & Vasander 1990: 22, as cited in Laine 1995: 52*) may be indicated by an arrow pointing left: 📖 Laine & Vasander 1990: 22 ⇐ 📖 Laine 1995: 52.

When the users are referring to private consultations, it may be necessary to distinguish the cases in which *new* information is provided by the consulting person from those cases in which the consultant *confirms* existing information. For example, in the *Finnish-Russian Forestry Dictionary* (Suomalais-venäläinen metsäsanakirja 2008), newly coined Russian equivalents suggested by the consultants were marked with a pen icon (✐) followed by the name of the consultant. Where a consultant *confirmed* the existence of a term, a checked tick box icon (☑) was used instead.

5.1.4 Basic requirements for bibliographic records

Description of bibliographic sources used in the compilation of a terminological collection should provide users with an answer to two vital questions:

- How reliable are the sources?
- How can they be accessed for additional information if necessary?

Bibliographic records also have to be accurate, uniform and compliant with one of the common bibliographic formats. The basic template for the documentation of written sources is provided in *Appendix 4.1*.

The most important data category in a basic bibliographic record is *bibliographic citation*. This category corresponds to the *bibliographicCitation* element in the inventory of metadata terms published by the *Dublin Core Metadata Initiative* (http://purl.org/dc/terms/bibliographicCitation, accessed 22.2.2013).

In the *bibliographic citation* field, information about the source should be provided in the way the contributor wants it to appear in the bibliography. There are no formal restrictions concerning the style for presenting bibliographic data, but it is recommended that contributors should follow a generally acknowledged bibliographic standard and be consistent.

Information provided about the source in the *bibliographic citation* field must be sufficient for its unambiguous identification. The following types of data are typically used for resource identification: name of the resource, responsibility (such as authors and responsible organization), place and date of publication (for published sources) or creation (for unpublished ones). Moreover, a reference to the host document and location within it has to be specified for contributions and information on the version or edition for repetitive works. The practical representation of these categories depends on the resource type.

Two other mandatory parameters which are not necessarily required for the identification of the resource but which are important pragmatically include *format of*

resource and *accessibility*. Values for the *format of resource* field should be picked from a predefined list. The following values borrowed from the *International Standard Bibliographic Description* (ISBD 2007: 0-1) will presumably cover the vast majority of cases: printed text, electronic resource, multimedia resource, moving images, sound recording, still images, cartographic resource, and notated music resource. However, it should also be possible to choose an *other source* option and specify it.

The *accessibility* data category informs the users of whether the source is public or private, published or unpublished, how it can be accessed and when it was retrieved. The distinction between public/private and published/unpublished sources may seem questionable at first glance, but it was established for a reason. Unpublished works are not necessarily private. For example, theses are usually unpublished but public resources. At the same time, access to published resources may be restricted.

If contributors want to provide additional important information about a source that does not fall into any of the above-mentioned categories, they should use the *supplementary information* field. It is also advisable to provide a reference to the bibliographic standard used for documenting the resource and indicate whether the bibliographic record is fully or partially compliant with it.

Sources in different scripts usually have to be placed in different sections of the bibliography (e.g., first Latin, then Cyrillic, then Chinese). The correct sorting of the sources requires that information be provided about the language and script of every bibliographic record.

A persistent *source ID* must be automatically assigned to every bibliographical record for its unambiguous identification. Each bibliographic record must also be provided with a unique visual source identifier that will serve as a reference to the bibliographic record in the entries. It is desirable that visual source identifiers should provide some information about the source, for example, a combination of the author and the year of publication. Documents created by several authors or with no author specified may be abbreviated by the initial letters of the title in uppercase, as suggested in the ISO standard (ISO 12615:2004: 8). The system must prevent

duplication of the visual source identifiers and may be programmed to suggest them automatically.

Since inline formatting (e.g., italics) is very common in bibliographic descriptions, the terminology management system has to provide means for adding inline formatting to the bibliographic records. Inline formatting should not have a negative impact on the functionality or speed of search.

In some cases, it may be necessary to contact the person who is responsible for the bibliographic record; for example, to request more information on the source or report mistakes. It is also useful to know who created and updated the record and when. This implies that bibliographic records should be provided with administrative data.

The proposed administrative data categories include c*reator, updater(s), creation date, update date(s)* and *responsible person(s)*. By default, the creator of the record becomes the responsible person unless specified otherwise. Categories other than *responsible person* should be automatically managed by the system and be read-only. The *responsiblePerson* category should contain a link to the person's profile or another record with contact information rather than a plain name. It is strongly recommended that a history of transactions related to bibliographic records should be kept and means for data restoration provided.

5.1.5   Documentation of private sources and contributors' expertise

By private sources, we mean people who give consultations on a private basis. In terminology work, these are usually domain experts and language specialists. When documenting private sources, it is important to take into account two recommendations provided in the ISO standard (ISO 12615:2004: 13). First, it is desirable to distinguish between information needed for the bibliographic reference and information required for contacting the person concerned. Second, in some countries there may be a legal requirement to declare that a file containing personal

data is being maintained. We provide a template for the documentation of private sources for bibliographic purposes in *Appendix 4.3*, and for contact purposes in *Appendix 4.4*. The latter template is largely based on the vCard standard (2009) but contains some modifications and extensions.

In a collaborative platform, contributors may also be viewed as private sources. However, their level of expertise needs to be documented in much more detail than is done for simply bibliographic purposes. We provide a template for the documentation of contributors' expertise in *Appendix 5*, which should be filled in by every new contributor upon registration to help evaluate the proficiency of contributors and the reliability of the data produced by them. Besides, it can be used for seeking people with a suitable background for new projects, administrative tasks, and so on.

In the template, contributors have to provide information about various kinds of expertise: general language, special domain, LSP, terminology work and collaborative content creation. A distinction is also made between various types of experience (e.g., obtained from education, scientific activities and practical work) and types of activity (such as technical writing, translating and teaching).

5.1.6 Advanced support for source management

Several features are described below which in our opinion should be implemented in the first place if designers of a collaborative terminology management system want to go beyond the minimal requirements and provide better support for the documentation of sources.

First, it is convenient for the users that bibliography may be automatically generated for any terminological collection or user-defined subset. Since users may have different preferences concerning the presentation style of bibliographic records, they should be able to add alternative versions of the *bibliographic citation* field. Users may then select the style in which they want bibliographic items to appear in

the bibliography. If a particular style is not available for all the sources included in the bibliography, the default version may be used and highlighted.

The minimum set of data may be extended and split into smaller data categories. There are plenty of bibliographic formats from which additional categories may be borrowed. We provide an extended template for the documentation of written sources in *Appendix 4.2*. The template is based on authoritative standards used for description and exchange of bibliographic data (such as ISBD 2007/2011; ISO 12615:2004; Functional Requirements for Bibliographic Records 2009; ISO 690:1987; ISO 690-2:1997; ISO 30042:2008; Betz & Schmitz 1999). Extended templates may be only used *in addition* to the basic template, not *instead* of it.

The template library may be enriched collaboratively. Other collaborative elements may include the feedback to the person responsible for a particular bibliographic record and forum-style comments on the source, as well as collaborative evaluation of sources through voting and ranking.

5.2 Administrative data

5.2.1 Definition of administrative data

The concept of *administrative data* is not defined explicitly in ISO terminological standards such as ISO 1087-2:2000, ISO 16642:2003, ISO 12620:1999, ISO 12620:2009 or ISO 30042:2008. In the 'Understanding Metadata' guide by the American National Standards Organization, administrative metadata is said to provide 'information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it' (NISO 2004: 1). In the context of terminology management, this definition could be modified in the following way: data related to terminology management entities that helps manage them.

Administrative data does not relate to terminological descriptions but to the data containers in which they are encapsulated, from the smallest units (data fields) all the way to the largest one (term bank). Administrative data should not be mixed with *structural metadata* that comprises semantic, syntactical and value domain specifications of *data categories* and other data classes. While data categories are abstractions, terminology management entities are their concrete manifestation in a terminological collection.

5.2.2 The potential multifunctionality of administrative data categories

Data categories may have several functions at once; for example, language indicators are an indispensable part of terminological description but they may also function as administrative data when they help to choose the correct spell-checker. On the other hand, such primarily administrative categories as *creator* and *update date* provide information that may help evaluate the quality and freshness of the data.

In other words, beyond the context of a particular application it is hard to determine whether data is administrative.

The potential multifunctionality of data should be taken into account when choosing or designing a schema for a new terminological collection. Syntactical restrictions imposed on administrative data are usually stricter than those imposed on terminological descriptions. If a particular data category in a terminological collection is used for both descriptive and administrative purposes or is anticipated to have administrative functions in the future, its specification as regards the value domain and datatype should be as strict as possible. If imposing direct restrictions on the value domain is undesirable, users may be forced to map user-created values onto a standardized list of permitted values. Another way to check and enforce the correct format of the data is to use validation tools.

Users should also be prevented from accidentally filling in administrative data fields instead of fields related to terminological descriptions. For example, a label like *obsolete* or *deprecated* may indicate that a particular LSP designation is outdated. At the same time, it may be used as an administrative marker indicating the system that a particular entity is no longer needed. Since the consequences of accidental or deliberate misuse of administrative fields can be dramatic, preventive measures have to be taken against it.

5.2.3 Administrative data categories in ISO 12620

ISO 12620:1999 contains an extensive list of administrative data categories, including the following upper-level data categories:
- terminology management transactions
- terminology management functions
- subset identifier
- authorization information
- user suggestion

- administrative term qualifiers
- language symbol
- foreign text
- collating sequence
- entry type
- element working status
- target database
- entry source
- concept identifier
- entry identifier
- record identifier
- file identifier
- cross-reference
- source
- source identifier
- namespace identifier
- origination entry.

Since administrative data categories tend to be system-specific, there is probably no use in commenting on or trying to supplement this list. Instead, we would like to suggest a classification of administrative data that may help plan data management in terminological resources.

5.2.4 Classification of administrative data

The classification of administrative data proposed below is multidimensional, i.e., a particular data category may belong to several upper classes representing various dimensions. The list of possible bases of division of administrative data is of course open. We have included only those that we found particularly useful for planning data

management in a terminology management system. A compact version of the classification can be found in *Appendix 2*.

Administrative data by the type of entity to which it relates

Terminology management entities vary from one data model to another, and administrative data may be classified by the type of entity to which it relates. The following terminology management entities (starting from the smallest) are typically distinguished in terminological resources (ISO 1087-2:2000; ISO 16642:2003; ISO 30042:2008; Wright 2001b):

- *Data field* – a container reserved for a particular *data element* – a unit of data that is considered indivisible in a certain context.
- *Term component section* – a part of a term section providing linguistic information about the components of a term.
- *Term section* – a part of a language section providing information about the term.
- *Language section* – a part of a terminological entry containing information related to one language.
- *Terminological entry* – a part of a terminological data collection that contains terminological data related to one concept.
- *Terminological data collection* – a collection of data containing information on concepts of specific subject fields as well as global and complementary information.
- *Term bank* – a compendium of terminological collections including the organizational framework for recording, processing and disseminating terminological data.

This list reflects the concept-oriented, hierarchical 'container inside container' data model optimized for storing data in XML documents and relational databases. In

ontology-based systems such as *TermFactory*, the list of terminology management entities may look different, but the differences between the hierarchical and relational viewpoints are largely reducible in practice.

Administrative data by the creator of the entity to which it relates

Most terminology management entities in a term bank are *predefined* in the data category schema. However, in a dynamic electronic environment, users have an opportunity to form user-defined objects by making queries or manually extracting and combining the data. Users may thus form *selections* of fields, records, collections or other selections based on considerations that are hard or impossible to formalize.

An advanced terminology management system may support storing, sharing and managing such user-defined objects and allow assignment of administrative data to them. In this case, administrative data may be divided into data related to predefined entities and related to user-defined entities. User-defined entities and the data related to them or required for their restoration (e.g., a query) may be stored in user profiles or exported into an external file.

Administrative data by chronological status

From the chronological perspective, data in a term bank may be actual or archived. Likewise, administrative data may be divided into related to the actual data and related to archived data. This distinction may be important from the point of view of version management.

Administrative data by the type of addition

Administrative data may be added manually or generated automatically, usually being generated automatically. In order to eliminate human error and ensure the correctness and consistency of the data, manual addition of administrative data should be guided and validated whenever possible.

Administrative data by the place of storage

Administrative data stored in the system may be called *internal* and data stored outside the system (e.g., in export files and backups) *external*. This division may be important from the point of view of version management, as well as data export and import. The same basis of division is applicable to terminology management entities. Administrative data stored within the entities is called *embedded* (cf. NISO 2004: 1).

Administrative data by access level

Some types of administrative data may be useful to ordinary users of a terminology management system and should be available to them either by default or on demand. Other types of administrative data are not intended for such users and should be available only to authorized users. Correspondingly, administrative data may be classified according to security access level.

Administrative data by its function

Administrative data may also be classified according to its function. The two most common types are data required for performing transactions and data about transactions.

Administrative data required for performing transactions

Terminology management systems are designed to perform operations on the data called *transactions*. Among the examples of administrative data required for performing a transaction are information about the object of the transaction (e.g., its identifier) and authorization information (e.g., which users or user groups and under what circumstances are authorized to perform the transaction).

Administrative data about transactions

The ability to keep track of various transactions and produce logs of the results of these transactions is an important feature of a terminology management system. Information about transactions on a terminology management entity includes at least the following types of data:

- type of transaction
- time of transaction
- initiator of transaction
- results of transaction.

Administrative data specifying the type of transaction

The most common transactions in terminology management are the following:
- *Input* – insertion of data into a data management system via its user interface.
- *Import* – incorporation of data from an external file or database, with or without prior modification.
- *Access* – retrieval of data from the system for various purposes.
- *Printing* – reproduction of data in printed form.
- *Export* – copying of data into an external file, with or without prior modification, for the purposes of use in another program.
- *Modification* – any changes made to data. Modification may be further divided into addition, deletion, substitution, merging, and so on.
- *Rollback* – an operation that returns data to some previous state.
- *Backup* – making copies of data so that these additional copies may be used to restore the original after a data loss event.
- *Archiving* – saving of historical records, with or without compression of the data.
- *Recovery* – the process of salvaging data from damaged, failed, corrupted, or inaccessible storage media when it cannot be accessed normally.
- *Securing* – protection of data from unauthorized access, theft or corruption while allowing data to remain accessible to its intended users.

Administrative data specifying the time of the transaction

Timestamps are widely used in terminological databases for marking the date and time of transactions. Various time and date formats exist, which may cause problems during import, export, backup and restoration of the data. To prevent this, the time and date format used in the terminological collection or terminology management system should always be explicitly specified (cf. Wright 2001a: 556).

We recommend using the ISO 8601:2004 extended date and time format *YYYY-MM-DD hh:mm:ss.* This format has been adopted by many language industry standards, such as TEI (Burnard & Bauman 2012) and TBX (ISO 30042:2008), and multiple relevant software applications, such as major database management systems (in particular, MySQL).

Time-zone corrections shall be specified as offsets from UTC (Coordinated Universal Time). The *offset* from UTC shall be given in the format ±[hh]:[mm]. For example, if the time being described is one hour ahead of UTC, the offset would be '+01:00'. The offset from UTC changes with daylight saving time (see ISO 8601:2004 or http://en.wikipedia.org/wiki/ISO_8601:2004, accessed 8.3.2013, for more details).

Administrative data specifying the initiator of the transaction

The initiator of a transaction may be a term bank user or the system itself (e.g., in the case of automatic backups). This information is important; for example, for securing the data and debugging.

Administrative data specifying the results of the transaction

This type of administrative data specifies whether a transaction was successful or unsuccessful and to what consequences it has led. If a transaction was unsuccessful, probable reasons and other debugging information may be provided. Otherwise, a report may be generated containing information on the objects that were affected by the transaction and the changes resulting from it.

6   Conclusion

The general trends and desiderata in the development of electronic reference products, including term banks, are clear: towards more content, more flexibility and customization, more user-friendliness and interactivity, better access and more connectivity with other sources of information (cf. Sobkowiak 2002; Atkins 2002; de Schryver 2003; Verlinde, Leroyer & Binon 2010). With the 'old' challenges of catering for human users better remaining in force, terminography and terminology management are also confronted with a number of new challenges, two of which we would like to highlight.

Large-scale collaborative content creation

While previous generations of term banks were 'read-only' resources, the content of which was in the hands of small groups of professional terminologists, many term banks have recently opened themselves up to interaction with the community of users and other stakeholders. Besides, in the past decade, a new type of term bank has emerged which with certain reservations can be called 'national term banks'.

At present, most national term banks are at the stage of collecting existing glossaries into one place, but in the future they could be much more. Since terminology is a huge part of lexis in modern languages and its most dynamic part, the scientific and societal impact of a well-designed national term bank is comparable to that of an academic dictionary, a national corpus, or a national WordNet project. The mission of a national term bank is to preserve and maintain the scientific and linguistic heritage and identity of the given language community and to provide a comprehensive and objective picture of the national conceptual and terminological landscape.

We believe that a truly national term bank can only be implemented as a collaborative project, allowing all major stakeholders to present their views on the terminology in their domain. We also assume that national term banks will gradually evolve into terminological 'stock exchanges' where terms are not only stored but also discussed, evaluated and validated by various stakeholders. Collaborative content creation and validation also seem to be the only realistic way to achieve an optimal balance between the quantity and quality of terminological data at affordable cost.

Collaborative terminology work may also be done at the domain and corporate level. The technology enabling real-time collaborative terminology work has been available for some time now (e.g., http://www.termwiki.com), and more solutions are being developed, including *TermFactory* and the platform declared in the FP7 project *TaaS – Terminology as a Service*. However, technology in itself is only one part of the puzzle. The collaborative approach presents multiple challenges related to the quality and credibility of collaboratively created content, prioritizing and counterbalancing of competing definitions and terminology, creation of an efficient and motivating ecosystem, as well as many others.

While there is a lot of research available on Wikipedia and some other successful content creation systems, the results of this research may not be directly projected onto collaborative terminology work, which has many peculiarities as opposed to other types of content creation. Our guide is one of the first contributions to the theory and methodology of collaborative terminology work but much more research and piloting is needed if the goal is to obtain high-quality terminological content at the end. Well-formulated principles are mentioned among the major reasons accounting for the success of Wikipedia (Lih 2004; Sanger 2005). Without proper methodology, quality assurance tools and professional guidance and moderation, the whole idea of collaborative terminology work may easily be compromised.

Data reusability and customization

In order to be cost-effective, modern electronic reference products have to be designed to serve both human users and computer agents. Terminology is of the utmost importance for NLP applications and other intelligent computer agents, such as machine translation systems, speech recognition and generation systems, computer assistants, and so on. Taking these systems into account requires that terminological descriptions should be comprehensive, detailed, accurate, unambiguous and presented in a machine-processable format.

A natural choice for presenting multipurpose data is ontologies, which have become a significant format for describing complex concept systems in many areas, especially in natural sciences and medicine. As terminology and ontology work are both rooted in concept analysis, combining them is a very logical and resource-saving step. At least three different (even partly opposed) approaches to uniting terminology and ontologies have been developed and three different terms coined: termontography (Temmerman & Kerremans 2003), ontoterminology (Roche et al. 2009) and term ontologies (Kudashev & Carlson 2012).

The benefits of using ontologies are not limited to the fact that they are a strict and machine-processable format. Ontologies are very flexible, provide excellent opportunities for dynamic and customizable presentation of content and are capable of bridging the gap between various metamodels, such as the concept-oriented and lemma-oriented models used in terminography and lexicography respectively.

Using an ontology-based data model instead of strictly hierarchical ones (e.g., plain XML) means much more than simply switching to a new technology. It is a paradigmatic shift from the linear, entry-centred dictionaries and vocabularies originating from the paper era to dynamic and customizable reference resources in

which users are able to freely choose the object, type and style of description according to their permanent or ad hoc needs. In our opinion, exploration of the potential of ontology-based data models may lead to the long-expected qualitative leap from today's 'computerized dictionaries' to an electronic dictionary of the Third Millennium (see Tarp 2009: 28–29, Tarp 2008: 124; de Schryver 2003: 188; Spohr 2008; Verlinde, Leroyer & Binon 2010).

However, in addition to technology-oriented development projects, fundamental theoretical research is required for achieving this goal. Unfortunately, decision-makers in academic circles do not seem to understand this, still considering terminology, terminography and lexicography as merely applied fields. As a result, specialized lexicography has only exploited the opportunities provided by the new media and technology to a limited degree, and there is an astonishing disproportion between practical needs for terminology and the state of research in the field (Tarp 2012 – 'Specialized Lexicography: 20 years in slow motion'). Something has to change here.

Bibliography

Adler, B.T., Chatterjee, K., de Alfaro, L., Faella, M., Pye, I. and Raman, V. (2008). Assigning Trust to Wikipedia Content. In *Proceedings of the 2008 International Symposium on Wikis (Porto, Portugal)*. New York: ACM. Available at http://users.soe.ucsc.edu/~luca/papers/08/ucsc-soe-08-07.pdf, accessed 19.11.2012.

Anthony, D., Smith, S.W. and Williamson, T. (2005). Explaining Quality in Internet Collective Goods: Zealots and Good Samaritans in the Case of Wikipedia. Paper presented in IANDE seminar, MIT. Available at http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf, accessed 19.11.2012.

Atkins, B.T.S. (2002). Bilingual Dictionaries: Past, Present and Future. In Corréard, M.-H. (ed.). *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. [S. l.]: EURALEX. 1–29.

Bargmeyer, B.E. and Gillman, D.W. (2000). Metadata Standards and Metadata Registries: An Overview. Paper presented at the International Conference on Establishment Surveys II, Buffalo, New York (2000). Available at http://www.bls.gov/ore/pdf/st000010.pdf, accessed 13.11.2012.

Bereznikova R.E. (1976). Podača nomenov v slovarâh raznyh tipov. In *Problematika opredelenij terminov v slovarâh raznyh tipov*. Leningrad: Nauka. 83–91.

Bergenholtz, H. and Tarp, S. (eds) (1995). *Manual of Specialised Lexicography. The Preparation of Specialised Dictionaries*. Amsterdam: Benjamins.

Berkov V.P. (2004). *Dvuâzyčnaâ leksikografiâ: Učebnik*. 2nd expanded ed. Moskva: Astrel' [etc.].

Betz, A. and Schmitz, K.-D. (1999). The Terminology Documentation Interchange Format TeDIF. In Sandrini, P. (ed.) *TKE '99: Terminology and Knowledge Engineering: Proceedings of the 5th International Congress on Terminology and Knowledge Engineering, 23–27 August 1999, Innsbruck, Austria*. Wien: TermNet. 782–792.

Bonono, R. (2000). Terminology for Translators – An Implementation of ISO 12620. *Meta*. Vol. 45, № 4. 646–669.

Burnard, L. and Bauman, S. (eds) (2012). *TEI P5: Guidelines for Electronic Text Encoding and Interchange by the TEI Consortium*. 2.2.0. Last updated on 25th October, 2012. Charlottesville, Virginia: Text Encoding Initiative Consortium. Available at http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf, accessed 30.11.2012.

Cabré, M.-T. (1999). *Terminology: Theory, Methods and Applications*. Amsterdam: Benjamins.

Cross, T. (2006). Puppy Smoothies: Improving the Reliability of Open, Collaborative Wikis. *First Monday*. Vol. 11, № 9. Available at http://pear.accc.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1400/1318, accessed 19.11.2012.

Davis, M. and Whistler, K. (2012). Unicode Collation Algorithm. Ver. 6.2.0. (2012-08-30). In *Unicode Technical Standard #10*. Unicode, Inc. Available at http://www.unicode.org/reports/tr10, accessed 30.11.2012.

Felber, H. (1984). *Terminology Manual* / by Helmut Felber [for] General Information Programme and UNISIST. Paris: UNESCO: International Information Centre for Terminology.

Functional Requirements for Bibliographic Records (2009) / International Federation of Library Associations and Institutions (IFLA), February 2009. Available at http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf, accessed 20.8.2009.

Geeraerts, D. Meaning and Definition. In Sterkenburg, P. van. (ed.) *A Practical Guide to Lexicography*. Amsterdam: Benjamins, 2003. 83–93.

Gerd, A.S. (1986). *Osnovy naučno-tehničeskoj leksikografii: (Kak rabotat' nad terminologičeskim slovarëm)*. Leningrad: Izdatel'stvo Leningradskogo universiteta.

Grinev, S.V. (1993). *Vvedenie v terminovedenie*. Moskva: Moskovskij Licej.

Grinev, S.V. (1995). *Vvedenie v terminografiû*. 2nd ed. Moskva: MPU.

Grinev-Grinevič, S.V. (2008). *Terminovedenie: Učebnoe posobie dlâ studentov vuzov*. Moskva: Akademiâ.

Halim, F., Yongzheng, W. and Yap, R. (2009). Wiki Credibility Enhancement. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (Orlando, Florida)*. New York: ACM. Available at http://www.wikisym.org/ws2009/procfiles/p117-halim.pdf, accessed 19.11.2012.

Handbook of Terminology (2001) / [compiled by] Silvia Pavel and Diane Nolet. Gatineau [Quebec, Canada]: Minister of Public Works and Government Services Canada, Terminology and Standardization Translation Bureau. Catalogue No. S53-28/2001. Available at http://www.btb.gc.ca/publications/documents/termino-eng.pdf, accessed 15.11.2012.

Handbook of Terminology Management (1997) / compiled by Sue Ellen Wright and Gerhard Budin. Vol. 1. Basic Aspects of Terminology Management. Amsterdam: Benjamins.

Handbook of Terminology Management (2001) / compiled by Sue Ellen Wright and Gerhard Budin. Vol. 2. Application-Oriented Terminology Management. Amsterdam: Benjamins.

Hartmann, R.R.K. (2001). *Teaching and Researching Lexicography*. New York: Longman.

Hartmann, R.K.K. and James, G. (1998). *Dictionary of Lexicography*. London; New York: Routledge.

Hutcheson, H. (2001). Practical Considerations for a Term Bank: Termium. In *Handbook of Terminology Management* / compiled by S.E. Wright and G. Budin. Vol. 2. Application-Oriented Terminology Management. Amsterdam: Benjamins. 666–676.

IETR 2005 = *Identification of Existing Terminology Resources. EUROTERMBANK project. Deliverable 2.1, version I.I, 31/12/2005* / Raguz, M., Raupauch, I. and Schmitz, K.-D. Available at http://project.eurotermbank.com/uploads/D2.1%20Identification%20of%20existing%20terminology%20resources.pdf, accessed 20.8.2010.

ISBD 2007 = *International Standard Bibliographic Description (ISBD)* (2007) / International Federation of Library Associations and Institutions (IFLA). Available at http://www.ifla.org/files/cataloguing/isbd/isbd-cons_2007-en.pdf, accessed 20.8.2009.

ISBD 2011 = *ISBD: International Standard Bibliographic Description: Consolidated Edition* (2011) / Ed. by Standing Committee of the IFLA Cataloguing Section. Berlin/Munich: De Gruyter.

ISO 639-1:2002. Codes for the Representation of Names of Languages – Part 1: Alpha-2 Code. Geneva: ISO.

ISO 639-2:1999. Codes for the Representation of Names of Languages – Part 2: Alpha-3 Code. Geneva: ISO.

ISO 690:1987. Documentation – Bibliographic References – Content, Form and Structure. Geneva: ISO.

ISO 690-2:1997. Information and Documentation – Bibliographic References – Part 2: Electronic Documents and Parts thereof. Geneva: ISO.

ISO 704:2009. Terminology Work – Principles and Methods. Geneva: ISO.

ISO 1087-1:2000. Terminology Work – Vocabulary – Part 1: Theory and Application. Geneva: ISO.

ISO 1087-2:2000. Terminology Work – Vocabulary – Part 2: Computer Applications. Geneva: ISO.

ISO 3166-1:2006. Codes for the Representation of Names of Countries and their Subdivisions – Part 1: Country Codes. Geneva: ISO.

ISO 8402:1994. Quality Management and Quality Assurance – Vocabulary. Geneva: ISO.

ISO 8601:2004. Data Elements and Interchange Formats – Information Interchange – Representation of Dates and Times. Geneva: ISO.

ISO 10241:1992. International Terminology Standards – Preparation and Layout. Geneva: ISO.

ISO 12199:2000. Alphabetical Ordering of Multilingual Terminological and Lexicographical Data Represented in the Latin Alphabet. Geneva: ISO.

ISO 12615:2004. Bibliographic References and Source Identifiers for Terminology Work. Geneva: ISO.

ISO 12616:2002. Translation-Oriented Terminography. Geneva: ISO.

ISO 12620:1999. Computer Applications in Terminology – Data Categories. Geneva: ISO.

ISO 12620:2009. Terminology and Other Language and Content Resources – Specification of Data Categories and Management of a Data Category Registry for Language Resources. Geneva: ISO.

ISO 15924:2004. Information and Documentation – Codes for the Representation of Names of Scripts. Geneva: ISO.

ISO 16642:2003 Computer Applications in Terminology – Terminological Markup Framework. Geneva: ISO.

ISO 24613:2008. Language Resource Management – Lexical Markup Framework (LMF). Geneva: ISO.

ISO 30042:2008. Systems to Manage Terminology, Knowledge and Content – TermBase eXchange (TBX). Geneva: ISO.

ISO/IEC 11179-1:2004. Information Technology – Metadata Registries (MDR) – Part 1: Framework. Geneva: ISO.

ISO/IEC 14651:2007. Information Technology. International String Ordering and Comparison – Method for Comparing Character Strings and Description of the Common Template Tailorable Ordering. Geneva: ISO.

Kandelaki, T.L. (1970). Rabota po uporâdočeniû naučhno-tehničeskoj terminologii i nekotorye lingvističeskie problemy, voznikaûŝie pri ètom. In *Lingvističeskie problemy naučno-tehničeskoj terminologii*. Moskva: Nauka. 40–52.

Kandelaki, T.L. (1973). K voprosu o nomenklaturnyh naimenovaniâh. In *Voprosy razrabotki naučno-tehničeskoj terminologii*. Riga: Zinatne. 60–71.

Kiâk, T.R. (1989). *Lingvističeskie aspecty terminovedeniâ: Učebnoe posobie dlâ studentov special'nosti 02.20*. Kiev: UMKVO.

Kudashev, I. (2005). On the Treatment of Nomenclature in Specialised Dictionaries. In Mård-Miettinen, K. and Niemelä, N. (eds) *Erikoiskielet ja käännösteoria. Vakki-symposiumi XXV. Vöyri 12.–13.2.2005*. Vaasa: Vaasan yliopisto, 2005. 176–183.

Kudashev, I. (2006). Additional Meaning Components of Terms and their Treatment in LSP Dictionaries. In Lehtinen, E. and Niemelä, N. (eds.) *Erikoiskielet ja käännösteoria. VAKKI-symposiumi XXVI. Vaasa 11.–12.2006*. Vaasa: Vaasan yliopisto, 2006. 143–149.

Kudashev, I. (2007) *Proektirovanie perevodčeskih slovarej special'noj leksiki*. Helsinki: Yliopistopaino. Also available at https://oa.doria.fi/bitstream/handle/10024/11997/proektir.pdf ?sequence=1, accessed 30.11.2012. (=Designing LSP Dictionaries for Translators. PhD thesis).

Kudashev, I. (2012). International Medical Classifications from the Terminological Point of View (Cases: International Classification of Diseases and Terminologia Anatomica). In Nissilä, N. and Siponkoski, N. (eds). *Kielet liikkeessä. VAKKI-symposiumi XXXII. Vaasa 10–11.2.2012*. Vaasa: University of Vaasa. 118–129.

Kudashev, I. and Carlson, L. Administrative Data in Terminological Collections and Term Ontologies. In Aguado de Cea et al. (eds). *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012), 19–22 June 2012, Madrid, Spain*. Madrid: TKE organizers. 270–282.

Kudashev, I. and Kudasheva, I. (2006). Software Demo: The Terminographic Processor MyTerMS. In Gilles-Maurice de Schryver (ed). *DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writing Systems, Tuesday 5th September 2006, Turin, Italy (Pre-EURALEX 2006)*. Pretoria: (SF)[2] Press, 35–40.

Kudashev, I. and Kudasheva, I. (2010). Semiotic Triangle Revisited for the Purposes of Ontology-based Terminology Management. In *TOTh 2010. Actes de la quatrième conférence TOTh – Annecy – 3 & 4 juin 2010*. Annecy: Institut Porphyre. 83–100.

Kudasheva, I. (2009). Terminologisen työn vaiheet laadittaessa Suomalais-venäläistä metsäsanakirjaa. In *Puusta katsoen. Metsätermit ja metsäsanakirjan laadinta* / Ed. by Inkeri Vehmas-Lehto. Helsinki: Helsingin yliopisto, Käännöstieteen laitos. 112–124.

Kudasheva, I. and Kudashev, I. (2008). *EU-Russia Project Co-operation Glossary (English-Finnish-Russian)*. Kouvola: University of Helsinki, Palmenia Centre for Continuing Education. Available at http://www.projectglossary.eu/download/rcg_v6.pdf, accessed 30.11. 2012.

Lampe, C. and Resnick, P. (2004). Slash(dot) and Burn: Distributed moderation in a large online conversation space. In *Conference on Human Factors in Computing Systems (CHI), Vienna, Austria, April 24–29, 2004*. New York: ACM. 543–550.

Landau, S. (2001). *Dictionaries. The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Leičik, V.M. (1974). Nomenklatura – promežutočnoe zveno meždu terminami i sobstvennymi imenami. In *Voprosy terminologii i lingvističeskoj statistiki*. Voronež: Voronežskij tehnologičeskij institut. 13–24.

Leičik, V.M. (1975). Leksikografičeskaâ terminologičeskaâ deâtel'nost' (klassifikaciâ i metodika sozdania terminologičeskih slovarej. In *Terminologiâ bibliotečnogo dela: Sbornik naučnyh trudov* / I.M. Suslova (ed.). Moskva.

Leičik, V.M. (2006). *Terminovedenie: predmet, metody, struktura*. Moskva: KomKniga.

Leičik, V.M. (2009). *Terminovedenie: predmet, metody, struktura*. 4th ed. Moskva: Librokom.

Lih, A. (2004). Wikipedia as Participatory Journalism: Reliable Sources? Metrics for Evaluating Collaborative Media as a News Resource. In *Proceedings of 5th International Symposium on Online Journalism (April 16–17, 2004, Austin, Texas)*. [S.I.]. 309–315. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.117.9104&rep=rep1& type=pdf, accessed 19.11.2012.

Lingvističeskaâ koncepciâ (1989) = *Lingvističeskaâ koncepciâ terminologičeskogo banka dannyh mašinnogo fonda russkogo âzyka: (Proekt)* / Belozerov V.N., Gerd A.S., Gorodeckij B.U. i dr. Moskva: MGUUÂ.

McNeill, J. et al. (eds) (2012). *International Code of Nomenclature for algae, fungi, and plants (Melbourne Code), Adopted by the Eighteenth International Botanical Congress Melbourne, Australia, July 2011* (electronic ed.). Bratislava: International Association for Plant Taxonomy. Available at http://www.iapt-taxon.org/nomen/main.php, accessed 14.1. 2013.

Merritt, J.E. and Bossenbroek, B.I. (1997). Basic Resources for Assigning Chemical Names within the Field of Chemical Nomenclature. In *Handbook of Terminology Management* / compiled by S.E. Wright and G. Budin. Vol. 1. Basic Aspects of Terminology Management. Amsterdam: Benjamins. 218–242.

Miram G.È. (2001). *Perevodnye kartinki. Professiâ: perevodčik*. Kiev: Nika-Centr.

Monaci, S. (2009). Quality Assessment Process in Wikipedia's Vetrina: The Role of the Community's Policies and Rules. *Observatorio (OBS*) Journal*, 8/2009. 148–161.

Nedobity, W. (1988). Classification Systems for Terminological Databanks. In Picken, C. (ed) *Translating and the Computer 9. Proceedings of a conference…* 12–13 November 1987, CBI Conference Centre, Centre Point, London. London: Aslib. Available at http://www.mt-archive.info/Aslib-1987-Nedobity.pdf, accessed 4.4.2010.

Nilsson, H. (2012). 'Critical Mass – Critical Masses? On Harmonizing in a National Termbank'. Presentation at the workshop 'Data Consistency and Interoperability in Terminological Resources' (19.6.2012, Madrid) held in connection with the TKE 2012 conference. Available at http://dantermbank.cbs.dk/dtb_uk/the_dantermbank_project_ launches_a_new_website/download_tke_workshop_slides, accessed 10.1.2013.

NISO. Understanding Metadata (2004). Bethesda [USA]: NISO Press. Available at http://www.niso.org/publications/press/UnderstandingMetadata.pdf, accessed 12.11.2012.

Noy, N.F. and McGuinness, D.L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. [Stanford: Stanford University, Knowledge Systems Laboratory]. Available at http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf, accessed 13.3.2013.

Nykänen, O. (1999). Sanastoprojektin vaiheet. In *Toimikunnista termitalkoisiin. 25 vuotta sanastotyön asiantuntemusta*. Helsinki: Tekniikan Sanastokeskus. 62–71.

Pasanen, P. (2004). "Roskaisia", sanoi terminologi koneella poimituista termeistä. In Koskela, M. and Pilke, N. (eds) *Erikoiskielet ja käännösteoria. Vakki-symposiumi XXVI, Vaasa 7.–8.2.2004*. Vaasa: Vaasan yliopisto. 243–251.

Pasanen, P. (2009). Merenkulun turvallisuuden koetinkiviä: Terminologisen tiedon poiminta teksteistä. Helsinki: Yliopistopaino.

Phillips, A. and Davis, M. (eds.) (2006) *Tags for Identifying Languages. Best Current Practice 47. Request for Comments 4646. September 2006*. Available at http://www.rfc-editor.org/rfc/rfc4646.txt, accessed 20.8.2009.

Picht, H. and Draskau, J. (1985). *Terminology: An Introduction*. Guildford: University of Surrey.

Priedhorsky, R., Chen, J., Lam, S., Panciera, K., Terveen, L. and Riedl, J. (2007). Creating, Destroying and Restoring Value in Wikipedia. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work, Sanibel Island, Florida, USA*. New York: ACM. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.123. 7456&rep=rep1&type=pdf, accessed 19.11.2012.

Reformatskij A.A. (1961). Čto takoe termin i terminologiâ. In *Voprosy terminologii (Materialy Vsesouznogo terminologičeskogo sovešaniâ)*. Moskva: Izdatel'stvo AN SSSR. 46–62.

Riehle, D. (2006). How and Why Wikipedia Works: An Interview with Angela Beesley, Elisabeth Bauer, and Kizu Naoko. In *Proceedings of the International Symposium on Wikis*. New York: ACM. 3–8. Available at http://portal.acm.org/ft_gateway.cfm?id=1149456& type=pdf, accessed 19.11.2012.

Roche, C., Calberg-Challot, M., Damas, L. and Rouard, Ph. (2009). Ontoterminology: A New Paradigm for Terminology. In *Proceedings of KEOD, Madeira, October 5–8.10.2009*. Available at http://hal.archives-ouvertes.fr/docs/00/62/21/32/PDF/Proceedings_KEOD_2009_ 128_CR.pdf, accessed 15.8.2012.

Sager, J.C. (1990). *A Practical Course in Terminology Processing*. Amsterdam: Benjamins.

Šajkevič, A.Â. and Ubin I.I. (1988). Vzglâd perevodčika i naučnogo rabotnika na perevodnoj terminologičeskij slovar'. In *Teoriâ i praktika naučno-tehničeskoj leksikografii: Sbornik statej*. Moskva: Russkij âzuk. 112–118.

Sanastotyön käsikirja. Soveltavan terminologian periaatteet ja työmenetelmät (1989) / Tekniikan sanastokeskus. Helsinki: Suomen Standardoimisliitto.

Sanger, L. (2005). *The Early History of Nupedia and Wikipedia*. Part II. April 19, 2005. Available at http://features.slashdot.org/article.pl?sid=05/04/19/1746205&tid=95, accessed 8.2.2010.

Šelov, S.D. (2003). *Termin. Terminologičnost'. Terminologičeskie opredeleniâ*. Sankt-Peterburg: Filologičeskij fakul'tet SpbGU.

Šelov, S.D. and Leičik, V.M. (2007). *Nomenklaturnye naimenovaniâ kak klass naučno-tehničeskoj leksiki: sostav i funkcii*. Sankt-Peterburg: Sankt-Peterburgskij gosudarstvennyj universitet.

Šhelov, S.D. and Leičik, V.M. (2012). *Terminologiâ i professional'naâ leksika: sostav i funkčii: Učebnoe posobie*. Sankt-Peterburg: SpbGU.

Schryver, G.-M. de (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*. Vol. 16, № 2. 143–199.

Sobkowiak, W. (2002). What can be but is not (and why) in Learners' MRDs. *Teaching English with Technology*. 2(3) (May 2002). Available at http://iatefl.org.pl/call/j_article9.htm, accessed 25.9.2010.

Spohr, D. (2008). Requirements for the Design of Electronic Dictionaries and a Proposal for their Formalisation. In Bernal, E. and DeCesaris, J. (eds). *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Barcelona: Universitat Pompeu Fabra. 617–629.

Suomalais-venäläinen metsäsanakirja (2008) / Kudasheva, I. and Kudashev, I. (authors); Vehmas-Lehto, I. and Gerd, A. (eds). Helsinki: Metsäkustannus.

Suonuuti, H. (1997). Sanastotyö lyhyesti. *Terminfo* 2/1997. 15–16.

Suonuuti, H. (2001). *Guide to Terminology*. Helsinki: Tekniikan Sanastokeskus.

Tatarinov, V.A. (1996). *Teoriâ terminovedeniâ. Vol. 1. Teoriâ termina: Istoriâ i sovremennoe sostoânie*. Moskva: Moskovskij Licej.

Tarp, S. (2008). *Lexicography in the Borderland Between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.

Tarp, S. (2009). Beyond Lexicography: New Visions and Challenges in the Information Age. In Bergenholtz, H., Nielsen, S. and Tarp, S. (eds) *Lexicography at a Crossroads. Dictionaries and Encyclopedias Today, Lexicographic Tools Tomorrow*. Bern: Peter Lang. 17–32.

Tarp, S. (2012). Specialised Lexicography: 20 years in slow motion. *Iberica*. № 24. 117–128.

Tekom Study (2010) = Successful Terminology Management in Companies. Practical Tips and Guidelines: Basic Principles, Implementation, Cost-Benefit Analysis, System

Overview / Straub, D., Schmitz, K.-D. and Gesellschaft für Technische Kommunikation. [Stuttgart]: Stuttgart TC and more GmbH.

Temmerman, R. and Kerremans, K. (2003). Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description. In *Proceedings of CIL17*. Prague: Matfyzpress.

Ubin, I.I. (1992*). ÈVM i slovar': (Metodičeskoe posobie)*. Moskva: VCP.

Varantola, K. (2002). Use and Usability in Dictionaries: Common Sense and Context Sensibility? In Corréard, M.-H. (ed). *Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins*. [S. l.]: EURALEX. 30–44.

Vasil'eva, N.V. (1988). O slovarâh meždunarodnyh terminoelementov. In *Teoriâ i praktika naučno-tehničeskoj leksikografii: Sbornik statej*. Moskva: Russkij âzuk. 77–82.

vCard standard (2009). Version 3.0. Available at http://www.imc.org/pdi, accessed 20.8.2010.

Verlinde, S., Leroyer, P. and Binon, J. (2010). Search and You Will Find. From Stand-Alone Lexicographic Tools to User Driven Task and Problem-Oriented Multifunctional Leximats. *International Journal of Lexicography*. Vol. 23, № 1. 1–17.

Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G. and Smith, M. (2008). Finding Social Roles in Wikipedia. In *Proceedings of the 2011 iConference*. New York: ACM. 122–129.

Wilkinson, D. and Huberman, B. (2007). Cooperation and Quality in Wikipedia. In *Proceedings of the 2007 International Symposium on Wikis (Montreal, Quebec, Canada)*. New York: ACM. 157–164. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi= 10.1.1.88.4591&rep=rep1&type=pdf, accessed 19.11.2012.

Wright, S.E. (1997). Term Selection: The Initial Phase of Terminology Management. In *Handbook of Terminology Management* / compiled by S.E. Wright and G. Budin. Vol. 1. Basic Aspects of Terminology Management. Amsterdam: Benjamins. 13–23.

Wright, S.E. (2001a). Data Categories in Terminology Management. In *Handbook of Terminology Management* / compiled by S.E. Wright & G. Budin. Vol. 2. Application-Oriented Terminology Management. Amsterdam: Benjamins. 552–571.

Wright, S.E. (2001b). Terminology Management Entry Structures. In *Handbook of Terminology Management* / compiled by S.E. Wright & G. Budin. Vol. 2. Application-Oriented Terminology Management. Amsterdam: Benjamins. 572–599.

Wöhner, T. and Peters, R. (2009). Assessing the Quality of Wikipedia Articles with Lifecycle Based Metrics. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration (Orlando, Florida)*. New York: ACM. Available at http://www.wikisym.org/ws2009/procfiles/p116-woehner.pdf, accessed 19.11.2012.

Appendix 1   Classification of terminological data

Appendix 1.1   Linguistic classification of terminological data

Values in angle brackets are coded forms of data categories used in *Appendix 1.2* for the sake of brevity.

1 Data related to form. *<Form>*
1.1 Data related to type of designation by its form. *<Form: type>*
1.2 Data related to spelling. *<Form: spelling>*
1.3 Data related to pronunciation. *<Form: pronunciation>*
    1.3.1 Data related to word formation. *<Form: formation>*
    1.3.2 Data related to inflection. *<Form: inflection>*

2 Data related to meaning. *<Meaning>*
2.1 Data related to logical meaning. *<Meaning: logical>*
2.2 Data related to type of designation by its logical meaning. *<Meaning: type>*
2.3 Data related to induced meaning. *<Meaning: induced>*
2.4 Data related to degree of transparency of designation. *<Meaning: transparency>*

3 Data related to encyclopedic description. *<EncyclopedicDescription>*

4 Data related to usage. *<Usage>*
4.1 Data related to usage restrictions. *<Usage: restrictions>*
    4.1.1 Data related to language restrictions. *<Usage: restrictions: language>*
    4.1.2 Data related to domain restrictions. *<Usage: restrictions: domain>*
    4.1.3 Data related to geographical restrictions. *<Usage: restrictions: geographical>*
    4.1.4 Data related to chronological restrictions. *<Usage: restrictions: chronological>*
    4.1.5 Data related to scientific school or theory restrictions. *<Usage: restrictions: schoolOrTheory>*
    4.1.6 Data related to organizational restrictions. *<Usage: restrictions: organizational>*
    4.1.7 Data related to proprietary restrictions. *<Usage: restrictions: proprietary>*
    4.1.8 Data related to register restrictions. *<Usage: restrictions: register>*
    4.1.9 Data related to professional group restrictions. *<Usage: restrictions: professionalGroup>*
    4.1.10 Data related to compliance restrictions. *<Usage: restrictions: compliance>*
    4.1.11 Data related to combinatory restrictions. *<Usage: restrictions: combinatory>*
4.2 Data related to frequency of use. *<Usage: frequency>*

5 Data related to lexical and logical relations. *<Relations>*

5.1 Data related to lexical relations. *<Relations: lexical>*

    5.1.1 Data related to synonymous relations. *<Relations: lexical: synonymous>*

    5.1.2 Data related to antonymous relations. *<Relations: lexical: antonymous>*

    5.1.3 Data related to homonymous relations. *<Relations: lexical: homonymous>*

    5.1.4 Data related to paronymous relations. *<Relations: lexical: paronymous>*

5.2 Data related to logical relations *<Relations: logical>*

    5.2.1 Data related to generic relations. *<Relations: logical: generic>*

    5.2.2 Data related to partitive relations. *<Relations: logical: partitive>*

5.3 Data related to non-hierarchical ontological relations. *<Relations: non-hierarchicalOntological>*

5.4 Data related to cross-language equivalence relations. *<Relations: equivalence>*

6 Data related to origin and development. *<OriginAndDevelopment>*

7 Other types of data related to description of LSP designation. *<OtherTerm-RelatedData>*

Appendix 1.2   Mapping of ISO 12620:1999 data categories onto linguistic classification of terminological data

This Appendix provides a sample mapping of the ISO 12620:1999 data categories onto the bridging linguistic classification of terminological data. The first line of each block contains the code and name of an ISO data category. This may be followed by a note related to the ISO category itself or the mapping. The last line of the block contains one or more linguistic data categories onto which the ISO data category may be mapped. A question mark after the data category means that mapping is uncertain and depends on the interpretation and actual contents of the original data category.

For the sake of brevity, the names of linguistic data categories are provided in a short but transparent form. These can be decoded with the help of *Appendix 1.1*. The mapping of subcategories is sometimes omitted if they can be mapped in the exactly the same way as their parent category. In such cases, the subcategories are listed in the note related to the parent category.

A.1 term
Note: formally, this data category coincides with spelling. If *term* is opposed to other types of LSP designations (e.g., appellations or nomenclature), this data category also provides information on the type of designation by its logical meaning.
Linguistic classification: *<Form: spelling>*; *<Meaning: type>*?

A.2.1 term type
Note: this data category includes heterogeneous data. See subcategories for more information.
Linguistic classification: *<Form>*?; *<Meaning: type>*?; *<Usage: restrictions>*?; *<Relations: lexical: synonymous>*?; *<OriginAndDevelopment>*?

A.2.1.1 main entry term
Note: this data category indirectly reflects preference.
Linguistic classification: *<Usage: restrictions: compliance>*

A.2.1.2 synonym

Note: the primary function of this data category is to reflect synonymous relations. Where opposed to *A.2.1.1 main entry term*, this data category also indirectly reflects preference.

Linguistic classification: *<Relations: lexical: synonymous>*; *<Usage: restrictions: compliance>*?

A.2.1.3 quasi-synonym

Note: same as previous.

Linguistic classification: *<Relations: lexical: synonymous>*; *<Usage: restrictions: compliance>*?

A.2.1.4 international scientific term

Note: depending on the interpretation of this data category, it may provide information about compliance, language restrictions, origination and development, and type of designation by its logical meaning.

Linguistic classification: *<Usage: restrictions: compliance>*; *<Usage: restrictions: language>*?; *<OriginAndDevelopment>*?; *<Meaning: type>*?

A.2.1.5 common name

Note: a common name is a synonym of an international scientific term that is used in general discourse.

Linguistic classification: *<Usage: restrictions: register>; <Usage: restrictions: compliance>*; *<Usage: restrictions: professionalGroup>*?; *<Meaning: type>*

A.2.1.6 internationalism

Note: depending on the interpretation and actual contents of this data category, it may provide information related to language restrictions and word formation, as well as origin and development.

Linguistic classification: *<Usage: restrictions: language>*?; *<Form: formation>*?; *<OriginAndDevelopment>*?

A.2.1.7 full form

Linguistic classification: <Form: type>

A.2.1.8 abbreviated form of term

Note: depending on the actual contents, this data category and its subcategories may provide information about the type of designation by its form, and about word formation and pronunciation, as well as origin and development.

Linguistic classification: *<Form: type>*; *<Form: formation>*?; *<Form: pronunciation>*?; *<OriginAndDevelopment>*

A.2.1.8.1 abbreviation

Linguistic classification: *<Form: type>*; *<Form: formation>*; *<OriginAndDevelopment>*

A.2.1.8.2 short form of term

Linguistic classification: *<Form: type>*; *<Form: formation>*; *<OriginAndDevelopment>*

A.2.1.8.3 initialism

Linguistic classification: *<Form: type>*; *<Form: formation>*; *<Form: pronunciation>*; *<OriginAndDevelopment>*

A.2.1.8.4 acronym

Linguistic classification: *<Form: type>*; *<Form: formation>*; *<Form: pronunciation>*; *<OriginAndDevelopment>*

A.2.1.8.5 clipped term

Linguistic classification: *<Form: type>*; *<Form: formation>*; *<OriginAndDevelopment>*

A.2.1.9 variant

Note: depending on the type of variant, this data category may provide information about compliance, spelling, pronunciation, word formation as well as origin and development.

Linguistic classification: *<Usage: restrictions: compliance>*; *<Form: spelling>*?; *<Form: pronunciation>*?; *<Form: formation>*?; *<OriginAndDevelopment>*?

A.2.1.10 transliterated form

Linguistic classification: *<Form: type>*

A.2.1.11 transcribed form

Linguistic classification: *<Form: type>*

A.2.1.12 romanized form

Linguistic classification: *<Form: type>*

A.2.1.13 symbol

Linguistic classification: *<Form: type>*

A.2.1.14 formula

Linguistic classification: *<Form: type>*

A.2.1.15 equation

Note: equations are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.1.16 logical expression

Note: logical expressions are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.1.17 materials management categories

Note: material management categories are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.1.17.1 sku

Note: skus are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.1.17.2 part number

Note: part numbers are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.1.18 phraseological unit

Note: in ISO 12620:1999, this category is split into three subcategories: *collocation*, *set phrase* and *synonymous phrase*. Collocations cannot be counted as phraseological units as they do not correspond to the definition of phraseological unit provided in the standard. Besides, collocations cannot be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*; *<Usage: restrictions: combinatory>*?; *<Relations: lexical: synonymous>*?

A.2.1.18.1 collocation

Note: collocations cannot be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Usage: restrictions: combinatory>*

A.2.1.18.2 set phrase

Linguistic classification: <Form: type>

A.2.1.18.3 synonymous phrase

Linguistic classification: <Form: type>; *<Relations: lexical: synonymous>*

A.2.1.19 standard text

Note: standard texts are unlikely to be objects of description in a terminological collection.

Linguistic classification: not applicable or *<Form: type>*

A.2.2 grammar

Note: this data category and its subcategories (A.2.2.1 part of speech, A.2.2.2 grammatical gender, A.2.2.3 grammatical number, A.2.2.4 animacy A.2.2.1 noun class and A.2.2.6 adjective class) provide information related to word inflection.

Linguistic classification: *<Form: inflection>*

A.2.3 usage

Linguistic classification: *<Usage>*

A.2.3.1 usage note

Linguistic classification: *<Usage>*

A.2.3.2 geographical usage

Linguistic classification: *<Usage: restrictions: geographical>*

A.2.3.3 register

Linguistic classification: *<Usage: restrictions: register>*

A.2.3.4 frequency

Linguistic classification: *<Usage: frequency>*

A.2.3.5 temporal qualifier

Linguistic classification: *<Usage: restrictions: chronological>*

A.2.3.6 time restriction

Linguistic classification: *<Usage: restrictions: chronological>*

A.2.3.7 proprietary restriction

Linguistic classification: *<Usage: restrictions: proprietary>*

A.2.4 term formation

Note: In ISO 12620:1999, this data category includes *etymology* that is separated from *data related to word formation* in our linguistic classification. Because of this, mapping of this data category onto *<Form: formation>* is uncertain.

Linguistic classification: *<Form: formation>*?; *<OriginAndDevelopment>*

A.2.4.1 term provenance

Note: term provenance is a classification of terms according to the methodology employed in their creation (e.g., transdisciplinary borrowing, translingual borrowing, loan translation or

neologism). This data category provides information about word formation as well as origin and development.

Linguistic classification: *<Form: formation>; <OriginAndDevelopment>*

A.2.4.2 etymology

Linguistic classification: *<OriginAndDevelopment>*

A.2.5 pronunciation

Linguistic classification: *<Form: pronunciation>*

A.2.6 syllabification

Linguistic classification: *<Form: pronunciation>*

A.2.7 hyphenation

Linguistic classification: *<Form: spelling>*

A.2.8 morphology

Linguistic classification: *<Form: formation>*

A.2.8.1 morphological element

Linguistic classification: *<Form: formation>*

A.2.8.2 term element

Linguistic classification: *<Form: formation>*

A.2.9 term status

Linguistic classification: *<Usage: restrictions: compliance>*

A.2.10 degree of synonymy

Linguistic classification: *<Relations: lexical: synonymous>*

A.3 equivalence

Note: this data category and its subcategories (A.3.1 degree of equivalence, A.3.2 false friend, A.3.3 directionality, A.3.4 reliability code and A.3.5 transfer comment) can all be mapped onto the corresponding data category in the linguistic classification.

Linguistic classification: *<Relations: equivalence>*

A.4 subject field

Note: this data category and its subcategories (A.4.1 classification system and A.4.2 classificaion number) contain information related to domain restrictions.

Linguistic classification: *<Usage: restrictions: domain>*

A.5 concept-related description

Note: this data category primarily provides information about the logical concept. However, as examples and contexts are also included in this category in ISO 12620:1999, it may contain other types of data as well.

Linguistic classification: *<Meaning: logical>*; *<Meaning: induced>*?; *<EncyclopedicDescription>*?; *<Usage>*?; *<Form>*?; *<Relations: logical>*?; *<Relations: non-hierarchicalOntological>*?; *<OriginAndDevelopment>*?; *<OtherRelatedData>*?

A.5.1 definition

Linguistic classification: *<Meaning: logical>*

A.5.2 explanation

Note: depending on its contents, this data category may provide information about logical or induced meaning as well as encyclopedic information.

Linguistic classification: *<Meaning: logical>*?; *<Meaning: induced>*?; *<EncyclopedicDescription>*?

A.5.3 context

Note: depending on its contents, this data category may provide information about form, meaning and usage as well encyclopedic information.

Linguistic classification: <Form>?; *<Meaning>*?; *<Usage>*?; *<EncyclopedicDescription>*?

A.5.4 example

Note: depending on its contents, this data category may provide information about form, meaning and usage as well as encyclopedic information.

Linguistic classification: *<Usage>*?; *<Meaning>*?; *<Form>*?; *<EncyclopedicDescription>*?

A.5.5 nontextual illustrations

Note: this data category and its subcategories (A.5.5.1 figure, A.5.5.2 audio, A.5.5.3 video, A.5.5.4 table, A.5.5.5 other binary data) may illustrate virtually any kind of data. Information about the data format (video, audio, etc.) is administrative data.

Linguistic classification: not applicable or *<Form>*?; *<Meaning>*?; *<EncyclopedicDescription>*?; *<Usage>*?; *<Relations>*?; *<OriginAndDevelopment>*?; *<OtherRelatedData>*?

A.5.6 unit

Note: measurement unit is encyclopedic information.

Linguistic classification: *<EncyclopedicDescription>*

A.5.7 range

Note: the set of limits within which a quantity is measured. This is encyclopedic information.

Linguistic classification: *<EncyclopedicDescription>*

A.5.8 characteristic

Linguistic classification: *<Meaning: logical>*

A.6 concept relation

Note: this data category provides information about various types of logical and ontological relations.

Linguistic classification: *<Relations: logical>?; <Relations: non-hierarchicalOntological>?*

A.6.1 generic relation

Linguistic classification: *<Relations: logical: generic>*

A.6.2 partitive relation

Linguistic classification: *<Relations: logical: partitive>*

A.6.3 sequential relation

Note: this data category and its subcategories (A.6.3.1 temporal relation and A.6.3.2 spatial relation) provide information about non-hierarchical ontological relations.

Linguistic classification: *<Relations: non-hierarchicalOntological>*

A.6.4 associative relation

Linguistic classification: *<Relations: non-hierarchicalOntological>*

A.7 conceptual structures

Note: this data category and its subcategories provide information about logical and non-hierarchical ontological relations.

Linguistic classification: *<Relations: logical>?; <Relations: non-hierarchicalOntological>?*

A.7.1 concept system

Linguistic classification: *<Relations: logical>?; <Relations: non-hierarchicalOntological>?*

A.7.2 concept position

Linguistic classification: *<Relations: logical>?; <Relations: non-hierarchicalOntological>?*

A.7.2.1 broader concept

Linguistic classification: *<Relations: logical>*

A.7.2.2 superordinate concept

Linguistic classification: *<Relations: logical>*

A.7.2.3 subordinate concept

Linguistic classification: *<Relations: logical>*

A.7.2.4 coordinate concept

Linguistic classification: *<Relations: logical>*

A.7.2.5 related concept

Linguistic classification: *<Relations: non-hierarchicalOntological>*

A.8 note

Note: this data category may contain information of any kind.

Linguistic classification: *<Form>?; <Meaning>?; <EncyclopedicDescription>?; <Usage>?; <Relations>?; <OriginAndDevelopment>?; <OtherRelatedData>?*

A.9 documentary language

Note: this data category and its subcategories (A.9.1 thesaurus name, A.9.2 thesaurus descriptor, A.9.2.1 top term, A.9.2.2 broader term, A.9.2.3 narrower term, A.9.2.4 related term, A.9.3 nondescriptor, A.9.4 keyword and A.9.5 index heading) are used for information retrieval purposes in library science and documentation environments. They may represent a different, more technical and less precise view on terms and concepts than the corresponding LSP. In principle, this information is not a part of terminological description. However, if there is no contradiction between interpretations of the term in LSP and an information retrieval system, this set of data categories provides information about relations.

Linguistic classification: not applicable or *<Relations>*

A.10 administrative information

Note: several data categories listed in ISO 12620:1999 under the title of administrative data are actually a part of terminological description. These categories are described below.

A.10.7 language symbol

Linguistic classification: *<Usage: restrictions: language>*

A.10.18.3 inverted form

Linguistic classification: *<Form: spelling>*

A.10.18.4 permuted term

Linguistic classification: *<Form: spelling >*

A.10.18.5 homograph

Linguistic classification: *<Relations: homonymous>*

A.10.18.6 antonym

Linguistic classification: *<Relations: lexical: antonymous>*

A.10.20 source identifier

Note: this data category (together with the corresponding bibliographical description of the source) may indirectly provide information about term usage.

Functional classification: *<termUsage>*?

Appendix 2   Classification of administrative data

1 Administrative data by the type of entity to which it relates

2 Administrative data by the creator of the entity to which it relates
1.1 Administrative data related to predefined entities
1.2 Administrative data related to user-defined entities

3 Administrative data by chronological status
3.1 Administrative data related to actual data
3.2 Administrative data related to archived data

4 Administrative data by the type of addition
4.1 Automatically generated administrative data
4.2 Manually added administrative data

5 Administrative data by the place of storage
    5.1.1 Administrative data stored inside the terminology management system
    5.1.2 Administrative data stored outside the terminology management system
    5.2.1 Administrative data stored inside the entity to which it relates
    5.2.2 Administrative data stored outside the entity to which it relates

6 Administrative data by access level

7 Administrative data by its function
7.1 Administrative data required for performing transactions
    7.1.1 Administrative data specifying the object of transaction
    7.1.2 Authorization information
7.2 Administrative data about transactions
    7.2.1 Administrative data specifying the type of transaction
    7.2.2 Administrative data specifying the time of the transaction
    7.2.3 Administrative data specifying the initiator of the transaction
    7.2.4 Administrative data specifying the results of the transaction

Appendix 3   Domain classification

Appendix 3.1   Domain classification in Finnish

1. Yleistermit

2. Luokittelemattomat aihealueet
2.1. luokittelematon tiedonala
2.2. luokittelematon toimiala

3. Antropologia

4. Arkeologia

5. Arkkitehtuuri (<- Rakennustaide)
5.1. arkkitehtuurin historia
5.2. arkkitehtuurin teoria
5.3. maisema-arkkitehtuuri
5.4. rakennetekniikka
5.5. rakennus- ja asuntosuunnittelu
5.6. sisustussuunnittelu
5.7. yhdyskuntasuunnittelu (<- yhteiskuntasuunnittelu; aluesuunnittelu)

6. Biologia
6.1. anatomia
6.2. biofysiikka
6.3. biokemia
6.4. biologinen antropologia (<- fyysinen antropologia)
6.5. biotekniikka
6.6. eläintiede (<- zoologia)
6.7. etologia
6.8. evoluutiobiologia
6.9. fysiologia
6.10. histologia (<- kudosoppi)
6.11. kasvitiede (<- botaniikka)
6.12. kehitysbiologia
6.13. mikrobiologia
6.14. molekyylibiologia
6.15. paleontologia
6.16. perinnöllisyystiede (<- genetiikka)
6.17. sienitiede (<- mykologia)
6.18. solubiologia
6.19. ympäristöbiologia

7. Ekologia ja ympäristönsuojelu

8. Elintarviketeollisuus
8.1. einesten ja valmisruokien valmistus

8.2. elintarviketekniikka

8.3. eläinten ruokien valmistus

8.4. hedelmien, marjojen ja kasvisten jalostus ja säilöntä

8.5. homogenoitujen ravintovalmisteiden ja dieettiruokien valmistus

8.6. juomien valmistus (<- juomateollisuus)

8.7. kalan, äyriäisten ja nilviäisten jalostus ja säilöntä

8.8. kaakaon, suklaan ja makeisten valmistus

8.9. kasvi- ja eläinöljyjen ja -rasvojen valmistus

8.10. leipomotuotteiden ja muiden jauhotuotteiden valmistus

8.11. liha- ja siipikarjatuotteiden valmistus

8.12. maitotaloustuotteiden valmistus

8.13. mausteiden ja maustekastikkeiden valmistus

8.14. mylly- ja tärkkelystuotteiden valmistus

8.15. sokerin valmistus

8.16. teen ja kahvin valmistus

8.17. teurastus ja lihan säilyvyyskäsittely

9. Eläinlääketiede ja eläinlääkintäpalvelut

10. Eläintenhoito ja eläinsuojelu

11. Eläintiede (<- Zoologia)

11.1. araknologia (<- hämähäkkitutkimus)

11.2. eläinmaantiede

11.3. herpetologia

11.4. hyönteistiede (<- entomologia)

11.5. kalatiede (<- iktyologia)

11.6. lintutiede (<- ornitologia)

11.7. mammalogia

11.8. paleozoologia

11.9. protozoologia

12. Energia-ala

12.1. energiakauppa

12.2. energian käyttö

12.3. energian siirto ja jakelu

12.4. energian tuotanto

12.5. energian varastointi

12.6. energiapolitiikka

12.7. energiateknologia

13. Farmakologia, farmasia ja lääketeollisuus (<- Lääkeaineoppi, farmasia ja lääketeollisuus)

13.1. biofarmasia

13.2. farmakodynamiikka

13.3. farmakogenetiikka

13.4. farmakogenomiikka

13.5. farmakognosia

13.6. farmakokinetiikka

13.7. farmaseuttinen kemia

13.8. farmasian teknologia

13.9. kliininen farmakologia

13.10. lääkevalvonta
13.11. lääkkeiden valmistus

14. Filosofia
14.1. arvoteoria (<- arvo-oppi; arvofilosofia; aksiologia)
14.2. estetiikka
14.3. etiikka
14.4. filosofiset suuntaukset
14.5. historianfilosofia
14.6. kielifilosofia
14.7. logiikka
14.8. metafysiikka
14.9. mielenfilosofia
14.10. oikeusfilosofia
14.11. ontologia
14.12. poliittinen filosofia
14.13. tieteenfilosofia
14.14. tietoteoria (<- tieto-oppi; epistemologia)
14.15. uskonnonfilosofia
14.16. yhteiskuntafilosofia

15. Folkloristiikka

16. Fysiikka
16.1. akustiikka (<- äänioppi)
16.2. atomifysiikka
16.3. biofysiikka
16.4. geofysiikka
16.5. hiukkasfysiikka
16.6. kvanttifysiikka
16.7. mekaniikka
16.8. molekyylifysiikka
16.9. optiikka
16.10. sähködynamiikka
16.11. termodynamiikka (<- lämpöoppi)
16.12. ydinfysiikka

17. Geodesia, kartografia ja geomatiikka (<- Maanmittausoppi, karttaoppi ja geomatiikka)
17.1. fotogrammetria
17.2. geoinformatiikka
17.3. kartoitus
17.4. maanmittaus
17.5. topografia

18. Geologia
18.1. geofysiikka
18.2. geokemia
18.3. geomorfologia
18.4. hydrogeologia
18.5. mineralogia ja kidetiede (<- mineralogia ja kristallografia)
18.6. paleontologia

18.7. petrografia
18.8. petrologia
18.9. sedimentologia ja stratigrafia
18.10. seismologia
18.11. taloudellinen geologia
18.12. vulkanologia

19. Hallinto ja johtaminen

20. Historia
20.1. diplomatiikka
20.2. epigrafia (<- epigrafiikka)
20.3. etnohistoria
20.4. heraldiikka
20.5. historiallinen maantiede
20.6. genealogia
20.7. kulttuurihistoria
20.8. numismatiikka
20.9. oikeushistoria
20.10. paleografia
20.11. papyrologia
20.12. poliittinen historia
20.13. sinettitiede
20.14. sosiaalihistoria
20.15. sotahistoria
20.16. taloushistoria
20.17. taidehistoria
20.18. tieteenhistoria
20.19. uskontohistoria

21. Hydrologia (<- Vesitiede)

22. Insinööritieteet
22.1. automaatiotekniikka
22.2. biotekniikka
22.3. kemian tekniikka (<- kemiantekniikka; kemian prosessitekniikka)
22.4. konetekniikka
22.5. lääketieteellinen tekniikka
22.6. materiaalitekniikka (<- materiaalioppi)
22.7. mittaustekniikka
22.8. nanotiede ja nanoteknologia
22.9. prosessitekniikka
22.10. rakennustekniikka
22.11. sähkötekniikka
22.12. tietojenkäsittelytiede
22.13. tuotantotalous
22.14. ydintekniikka
22.15. ympäristötekniikka

23. Jalkineteollisuus (<- Jalkineiden valmistus; Kenkäteollisuus)

24. Julkishallinnon toiminta

25. Järjestöjen toiminta
25.1. ammattialajärjestöjen toiminta
25.2. ammattiyhdistysten toiminta
25.3. elinkeinoelämän ja työnantajajärjestöjen toiminta
25.4. kansainvälisten järjestöjen toiminta
25.5. kansalaisjärjestöjen toiminta
25.6. poliittisten järjestöjen toiminta
25.7. uskonnollisten järjestöjen toiminta

26. Kaivostoiminta ja louhinta
26.1. kiven louhinta
26.2. hiekan ja saven otto
26.3. kemiallisten ja lannoitemineraalien louhinta
26.4. kivihiilen ja ruskohiilen kaivu
26.5. metallimalmien louhinta
26.6. raakaöljyn ja maakaasun tuotanto
26.7. suolan tuotanto
26.7. turpeen nosto

27. Kalastus ja kalatalous
27.1. ammattikalastus
27.2. kalanjalostus
27.3. kalanviljely (<- vesiviljely; kalankasvatus)
27.4. ravustus ja raputalous
27.5. virkistyskalastus

28. Kansainväliset suhteet
28.1. diplomatia
28.2. kansainvälinen kauppa
28.3. kansainvälinen oikeus
28.4. kansainvälinen politiikka
28.5. kansainvälisten järjestöjen toiminta
28.6. ulkopolitiikka

29. Kansatiede ja etnografia (<- Etnologia ja etnografia)

30. Kasvatus ja koulutus
30.1. kasvatustiede
30.2. koulutusjärjestelmä
30.3. koulutuspalvelut
30.4. koulutuspolitiikka

31. Kasvitiede (<- Botaniikka)
31.1. ekonominen kasvitiede
31.2. kasvianatomia
31.3. kasviekologia
31.4. kasvifysiologia
31.5. kasvigenetiikka
31.6. kasvimaantiede (<- geobotaniikka)

31.7. kasvimorfologia
31.8. kasvisystematiikka (<- kasvitaksonomia)
31.9. paleobotaniikka

32. Kauppa ja talous
32.1. asiakaspalvelu
32.2. kaupankäynti
32.3. kirjanpito ja tilintarkastus
32.4. kulutus
32.5. markkinointi
32.6. rahoitus
32.7. rahoitustoiminta
32.8. taloushistoria
32.9. talousjärjestelmä
32.10. talousoikeus
32.11. talouspolitiikka
32.12. taloustiede
32.13. vakuutus
32.14. verotus
32.15. vuokraus- ja leasingtoiminta
32.16. yritystoiminta

33. Kemia
33.1. analyyttinen kemia
33.2. biokemia
33.3. epäorgaaninen kemia
33.4. fysikaalinen kemia
33.5. orgaaninen kemia

34. Kemianteollisuus
34.1. kemian tekniikka (<- kemiantekniikka; kemian prosessitekniikka; kemianteknologia; teknillinen kemia)
34.2. kemikaalien ja kemiallisten tuotteiden valmistus
34.3. koksin ja jalostettujen öljytuotteiden valmistus
34.4. kumituotteiden valmistus
34.5. lääkkeiden valmistus
34.6. muovituotteiden valmistus

35. Kielitiede (<- Lingvistiikka)
35.1. dialektologia
35.2. etnolingvistiikka
35.3. etymologia
35.4. fonetiikka
35.5. fonologia
35.6. funktionaalinen kielitiede
35.7. kielifilosofia
35.8. kielihistoria (<- historiallinen kielitiede)
35.9. kielimaantiede
35.10. kielioppi
35.11. kielipolitiikka

35.12. kieliteknologia
35.13. kielitypologia
35.14. kielten opetus
35.15. kognitiivinen lingvistiikka
35.16. käännöstiede
35.17. leksikologia
35.18. leksikografia
35.19. neurolingvistiikka
35.20. oikeuslingvistiikka
35.21. pragmatiikka
35.22. psykolingvistiikka
35.23. semantiikka
35.24. sosiolingvistiikka
35.25. stilistiikka (<- tyylioppi)
35.26. tekstilingvistiikka
35.27. terminologia
35.28. vertaileva kielitiede

36. Kiinteistöala
36.1. kiinteistöjohtaminen
36.2. kiinteistökauppa
36.3. kiinteistön- ja maisemanhoito
36.4. kiinteistönvälitys
36.5. kiinteistöjen vuokraus

37. Kirjallisuudentutkimus (<- Kirjallisuustiede)
37.1. kirjallisuudenhistoria
37.2. kirjallisuuskritiikki

38. Kosmetologia ja kauneudenhoitopalvelut

39. Kotitalous
39.1. kodinhoito
39.2. kodin taloustiede
39.3. taloustarvikkeet

40. Kulttuuri ja viihde
40.1. kulttuuri- ja viihdelaitosten toiminta
40.2. kulttuuri- ja viihdetapahtumien järjestäminen
40.3. kulttuuria ja viihdettä palveleva toiminta
40.4. kulttuuripolitiikka

41. Kuvataide
41.1. grafiikka
41.2. kalligrafia (<- kaunokirjoitus; kaunokirjoitustaide)
41.3. kuvanveisto (<- kuvanveistotaide)
41.4. maalaustaide
41.5. piirustustaide
41.6. taidehistoria
41.7. taidekritiikki
41.8. valokuvataide (<- valokuvaustaide)

42. Käyttötaide

42.1. keramiikkataide
42.2. kirjataide
42.3. korutaide
42.4. käyttögrafiikka (<- taideteollinen grafiikka)
42.5. lasitaide
42.6. metallitaide
42.7. paperitaide
42.8. puutaide
42.9. tekstiilitaide

43. Lasikeraaminen teollisuus

43.1. keraamisten tuotteiden valmistus
43.2. lasin ja lasituotteiden valmistus

44. Liikenne ja logistiikka

44.1. henkilöliikenne
44.2. huolinta
44.3. ilmaliikenne
44.4. avaruusliikenne
44.5. lastinkäsittely
44.6. liikennepolitiikka
44.7. liikenneturvallisuus
44.8. muuttopalvelut
44.9. navigointi
44.10. posti- ja kuriiripalvelut
44.11. putkijohtokuljetukset
44.12. rautatieliikenne
44.13. tavaraliikenne
44.14. tieliikenne
44.15. tullaus
44.16. varastointi
44.17. vesiliikenne

45. Liikunta ja urheilu

45.1. liikuntakasvatus
45.2. liikuntapalvelut
45.3. urheilulaitosten ja urheiluseurojen toiminta
45.4. urheilulajit
45.5. urheiluvälineet

46. Lääketiede ja terveydenhuolto

46.1. anatomia
46.2. anestesiologia
46.3. biolääketiede
46.4. ehkäisevä lääketiede (<- ennaltaehkäisevä lääketiede)
46.5. epidemiologia
46.6. endokrinologia
46.7. fysiatria
46.8. fysiologia

46.9. gerontologia
46.10. gynekologia (<- naisten taudit)
46.11. hammaslääketiede
46.12. hematologia
46.13. hepatologia
46.14. hoitomenetelmät
46.15. ihotautioppi
46.16. immunologia
46.17. infektiotaudit
46.18. kasvainoppi
46.19. katastrofilääketiede
46.20. keuhkotaudit (<- pulmonologia)
46.21. kirurgia
46.22. korva-, nenä- ja kurkkutautioppi
46.23. lastentautioppi
46.24. munuaisoppi
46.25. neurologia
46.26. oikeuslääketiede
46.27. patologia
46.28. perinnöllisyyslääketiede
46.29. psykiatria
46.30. radiologia
46.31. ravitsemustiede
46.32. reumaoppi
46.33. seksuaalilääketiede
46.34. silmätautioppi
46.35. sydäntautioppi
46.36. synnytysoppi
46.37. terveydenhuoltojärjestelmä
46.38. terveydenhuoltopalvelut
46.39. terveyspolitiikka
46.40. traumatologia
46.41. urologia
46.42. vaihtoehtoinen lääketiede
46.43. vatsatautioppi (<- gastroenterologia)

47. Maantiede

47.1. aluemaantiede
47.2. eläinmaantiede
47.3. geomorfologia
47.4. historiallinen maantiede
47.5. kasvimaantiede
47.6. kulttuurimaantiede
47.7. maantieteelliset nimet
47.8. paleomaantiede
47.9. poliittinen maantiede
47.10. talousmaantiede
47.11. yhteiskuntamaantiede

48. Maaperätiede (<- Maaperäoppi)

49. Maatalous
49.1. kotieläinhoito
49.2. maanviljely
49.3. maatalouspolitiikka
49.4. maataloustiede (<- agronomia)

50. Majoitus- ja ravitsemisala
50.1. elintarvikehygienia
50.2. gastronomia
50.3. majoitusliikkeiden toiminta
50.4. ravitsemisliikkeiden toiminta
50.5. ravitsemustiede
50.6. ruokakulttuuri

51. Matematiikka
51.1. algebra
51.2. analyysi
51.3. aritmetiikka (<- laskuoppi)
51.4. geometria
51.5. lukuteoria
51.6. todennäköisyysteoria
51.7. topologia

52. Matkailu (<- Turismi)
52.1. matkailumuodot
52.2. matkatoimistojen ja matkanjärjestäjien toiminta

53. Metalliteollisuus
53.1. koneteollisuus
53.2. metallituoteteollisuus

54. Metallien jalostus

55. Meteorologia ja klimatologia (<- Ilmatiede ja ilmastotiede)

56. Metrologia

57. Metsästys ja riistatalous

58. Metsätalous
58.1. metsien monikäyttö
58.2. metsäekologia
58.3. metsäekonomia
58.4. metsägenetiikka ja metsänjalostus
58.5. metsänhoito
58.6. metsänsuojelu
58.7. metsäntutkimus
58.8. metsäpolitiikka
58.9. metsäsuunnittelu
58.10. metsäteknologia

58.11. puunkorjuu
58.12. puun ja metsän mittaus

59. Metsäteollisuus
59.1. massa- ja paperiteollisuus
59.2. puuteollisuus

60. Muotoilu

61. Musiikki
61.1. musiikin ja äänitteiden tuotanto
61.2. musiikin tyylilajit
61.3. musiikkitiede (<- musikologia)

62. Nahkateollisuus
62.1. nahan valmistus
62.2. nahkatuotteiden valmistus
62.3. nahkavaatteiden valmistus

63. Oikeus ja lainsäädäntö
63.1. finanssioikeus
63.2. hallinto-oikeus
63.3. kansainvälinen oikeus
63.4. kauppaoikeus
63.5. lakiasiainpalvelut
63.6. lakien säätäminen
63.7. maaoikeus
63.8. oikeudet ja vapaudet
63.9. oikeusfilosofia
63.10. oikeushistoria
63.11. oikeusjärjestelmä
63.12. oikeustiede (<- juridiikka; lakitiede)
63.13. perheoikeus
63.14. prosessioikeus
63.15. rikosoikeus
63.16. sopimusoikeus
63.17. työoikeus
63.18. valtiosääntöoikeus
63.19. vero-oikeus
63.20. vesioikeus
63.21. ympäristöoikeus

64. Perinnöllisyystiede (<- Genetiikka)
64.1. evolutiivinen genetiikka
64.2. geenitekniikka
64.3. genomiikka (<- genomitutkimus)
64.4. ihmisgenetiikka
64.5. klassinen genetiikka
64.6. molekyyligenetiikka
64.7. populaatiogenetiikka
64.8. sytogenetiikka

65. Politiikka ja politiikan tutkimus

65.1. kansainvälinen politiikka
65.2. poliittinen filosofia
65.3. poliittinen historia
65.4. poliittinen ideologia
65.5. poliittinen järjestelmä
65.6. poliittinen taloustiede
65.7. poliittisten järjestöjen toiminta
65.8. politiikan sosiologia
65.9. sisäpolitiikka
65.10. ulkopolitiikka

66. Projektinhallinta

67. Psykologia
67.1. biopsykologia
67.2. evoluutiopsykologia
67.3. havaintopsykologia
67.4. kasvatuspsykologia
67.5. kehityspsykologia
67.6. kliininen psykologia
67.7. kognitiivinen psykologia
67.8. kokeellinen psykologia
67.9. kulttuuripsykologia
67.10. liikennepsykologia
67.11. neuropsykologia
67.12. oikeuspsykologia
67.13. persoonallisuuden psykologia (<- persoonallisuuspsykologia)
67.14. psykolingvistiikka
67.15. sosiaalipsykologia
67.16. työ- ja organisaatiopsykologia
67.17. urheilupsykologia
67.18. uskontopsykologia
67.19. ympäristöpsykologia

68. Puutarhasuunnittelu ja puutarhanhoito

69. Rahapeli- ja vedonlyöntitoiminta

70. Rakennusteollisuus ja rakentaminen
70.1. rakennusaineiden ja rakennustuotteiden valmistus
70.2. rakennusasennus
70.3. rakennus- ja asuntosuunnittelu
70.4. rakennuspaikan valmistelutyöt
70.5. rakennusten ja rakennelmien purku
70.6. rakennusten ja rakennelmien viimeistely
70.7. rakennustekniikka
70.8. rakennuttaminen

71. Semiotiikka

72. Sirkus ja muut viihdetaiteet
72.1. sirkustaide
72.2. stand up-komiikka
72.3. imitaatio

73. Sodankäynti ja maanpuolustus
73.1. aseet ja sotatekniikka
73.2. asepalvelus ja siviilipalvelus
73.3. asevoimien rakenne ja toiminta
73.4. puolustuspolitiikka
73.5. rauhanturvaaminen
73.6. siviilipuolustus
73.7. sotahistoria
73.8. sotaoikeus
73.9. sotataito
73.10. sotatalous
73.11. sotatiede
73.12. sotilashallinto
73.13. turvallisuuspolitiikka

74. Sosiaaliturva
74.1. sosiaalietuudet
74.2. sosiaalihuolto
74.3. sosiaalipalvelut
74.4. sosiaalipolitiikka
74.5. sosiaalivakuutus

75. Sosiologia
75.1. historiallinen sosiologia
75.2. kasvatussosiologia
75.3. kaupunkisosiologia
75.4. kulttuurisosiologia
75.5. kulutussosiologia
75.6. maaseutusosiologia
75.7. oikeussosiologia
75.8. perhesosiologia
75.9. politiikan sosiologia
75.10. sosiaalihistoria
75.11. sosiolingvistiikka
75.12. sotilassosiologia
75.13. taiteensosiologia
75.14. taloussosiologia
75.15. terveyssosiologia
75.16. tiedonsosiologia
75.17. tieteensosiologia
75.18. työnsosiologia (<- työsosiologia)
75.19. viestinnän sosiologia
75.20. uskontososiologia
75.21. ympäristösosiologia

76. Standardointi

77. Sähkö- ja elektroniikkateollisuus
77.1. elektronisten komponenttien valmistus
77.2. elektronisten lääkintä- ja terapialaitteiden valmistus
77.3. mittaus- ja testauslaitteiden valmistus
77.4. navigointilaitteiden valmistus
77.5. optisten instrumenttien valmistus
77.6. sähkölaitteiden valmistus
77.7. säteilylaitteiden valmistus
77.8. tallennevälineiden valmistus
77.9. tietokoneiden ja tietokoneiden oheislaitteiden valmistus
77.10. valokuvausvälineiden valmistus
77.11. viestintälaitteiden valmistus
77.12. viihde-elektroniikan valmistus

78. Tanssi
78.1. koreografia
78.2. tanssin tyylilajit

79. Teatteri ja elokuva-ala
79.1. lavastus
79.2. maskeeraus
79.3. näytteleminen
79.4. ohjaus
79.5. tarpeisto
79.6. teatteri- ja elokuvakritiikki
79.7. teatterin tyylilajit
79.8. teatteritiede
79.9. valosuunnittelu
79.10. videoiden ja elokuvien tuotanto
79.11. äänisuunnittelu

80. Tekniset palvelut
80.1. arkkitehtipalvelut
80.2. asennus-, korjaus- ja huoltopalvelut
80.3. insinööripalvelut
80.4. tekninen konsultointi
80.5. tekninen tarkastus
80.6. tekninen testaus ja analysointi

81. Tekstiiliteollisuus
81.1. kankaiden kudonta
81.2. tekstiilien viimeistely
81.3. tekstiilikuitujen valmistelu ja kehruu
81.4. tekstiilituotteiden valmistus
81.5. tekstiilivaatteiden valmistus

82. Tieteentutkimus
82.1. tieteenetiikka
82.2. tieteenfilosofia

82.3. tieteenhistoria
82.4. tieteensosiologia

83. Tietojen kokoaminen ja tallentaminen
83.1. arkistonhoito
83.2. arkistotiede
83.3. kirjastopalvelut
83.4. kirjastotiede
83.5. museoiden toiminta
83.6. museologia

84. Tietojenkäsittely ja tietotekniikka
84.1. tietojenkäsittelytiede
84.2. tieto- ja viestintätekniikka (<- tieto- ja viestintäteknologia; TVT)
84.3. tietotekniikkapalvelut

85. Tilastotiede (<- Statistiikka)

86. Toksikologia

87. Tupakkateollisuus

88. Turkisteollisuus
88.1. turkisten muokkaus ja värjäys
88.2. turkistuotteiden valmistus
88.3. turkisvaatteiden valmistus

89. Turvallisuusala (<- Turva-ala)
89.1. etsivätoiminta
89.2. henkilösuojelu
89.3. hälytyskeskustoiminta
89.4. järjestyksenvalvonta
89.5. liikenneturvallisuus
89.6. nuohous
89.7. palotoimi
89.8. pelastustoimi
89.9. poliisitoimi
89.10. rajavalvonta
89.11. riskienhallinta
89.12. säteilysuojelu
89.13. tiedustelupalvelut
89.14. turvatarkastustoiminta
89.15. turvallisuustekniikka (<- turvatekniikka)
89.16. työturvallisuus
89.17. vankeinhoito

90. Tutkimus ja kehittäminen
90.1. tutkimuksen ja kehittämisen infrastruktuuri
90.2. tutkimuspolitiikka
90.3. tutkimusta palveleva toiminta

91. Työelämä ja työsuojelu

91.1. ammatinvalinnanohjaus
91.2. ammattialajärjestöjen toiminta
91.3. ammattiyhdistysten toiminta
91.4. elinkeinoelämän ja työnantajajärjestöjen toiminta
91.5. ergonomia
91.6. henkilöstöjohtaminen
91.7. henkilöstökoulutus
91.8. henkilöstön hankinta
91.9. työllistämistoiminta
91.10. työn organisointi
91.11. työoikeus
91.12. työterveyshuolto
91.13. työturvallisuus
91.14. työvoimapolitiikka

92. Tähtitiede ja avaruustutkimus (<- Astronomia ja avaruustutkimus)

92.1. astrofysiikka
92.2. astrometria
92.3. kosmologia
92.4. taivaanmekaniikka

93. Uskomukset ja okkultismi

93.1. alkemia
93.2. antroposofia
93.3. astrologia
93.4. ennustaminen
93.5. fengshui
93.6. kabbala
93.7. magia
93.8. numerologia
93.9. parapsykologia
93.10. spiritualismi
93.11. teosofia
93.12. ufologia (<- ufotutkimus)

94. Uskonto

94.1. seurakuntien ja uskonnollisten järjestöjen toiminta
94.2. teologia
94.3. uskonnonfilosofia
94.4. uskonnot ja uskonnolliset suuntaukset
94.5. uskontohistoria
94.6. uskontokasvatus (<- uskonnollinen kasvatus)
94.7. uskontotiede

95. Vapaa-aika ja harrastukset

96. Vesi-, viemäri- ja jätehuolto

96.1. jätteen keruu
96.2. jätteen käsittely ja loppusijoitus

96.3. materiaalien kierrätys
96.4. veden otto, puhdistus ja jakelu
96.5. viemäri- ja jätevesihuolto

97. Viestintä ja tiedonvälitys

97.1. joukkoviestimien toiminta
97.2. journalismi ja journalistiikka
97.3. kustannus- ja julkaisutoiminta
97.4. painaminen ja tallenteiden jäljentäminen
97.5. radio- ja televisio-ohjelmien tuotanto
97.6. televiestintä (<- kaukoviestintä)
97.7. tietopalvelut
97.8. viestintäpolitiikka
97.9. viestintätieteet

98. Väestötiede (<- Demografia)

Appendix 3.2   Domain classification in English

1. General Terms

2. Unclassified domains
2.1. unclassified special knowledge field
2.2. unclassified activity field

3. Anthropology

4. Archaeology (<- Archeology)

5. Architecture
5.1. history of architecture
5.2. theory of architecture
5.3. landscape architecture
5.4. structural engineering
5.5. constructional planning and housing design
5.6. interior design
5.7. regional planning

6. Biology
6.1. anatomy
6.2. biophysics
6.3. biochemistry
6.4. biological anthropology (<- physical anthropology)
6.5. biotechnology
6.6. zoology
6.7. ethology
6.8. evolutionary biology
6.9. physiology
6.10. histology
6.11. botany
6.12. developmental biology
6.13. microbiology
6.14. molecular biology
6.15. palaeontology
6.16. genetics
6.17. mycology
6.18. cell biology
6.19. environmental biology

7. Ecology and environmental protection

8. Food industry
8.1. manufacture of prepared meals and dishes
8.2. food technology
8.3. manufacture of prepared animal feeds
8.4. processing and preserving of fruit and vegetables

8.5. manufacture of food preparations and dietetic food

8.6. manufacture of beverages

8.7. processing and preserving of fish, crustaceans and molluscs

8.8. manufacture of cocoa, chocolate and sugar confectionery

8.9. manufacture of vegetable and animal oils and fats

8.10. manufacture of bakery and farinaceous products

8.11. production of meat and poultry meat products

8.12. manufacture of dairy products

8.13. manufacture of condiments and seasonings

8.14. manufacture of grain mill products, starches and starch products

8.15. manufacture of sugar

8.16. processing of tea and coffee

8.17. processing and preserving of meat

9. Veterinary medicine and veterinary services

10. Animal care and protection of animals

11. Zoology
11.1. arachnology

11.2. animal geography (<- zoogeography)

11.3. herpetology

11.4. entomology

11.5. ichthyology

11.6. ornithology

11.7. mammalogy

11.8. paleozoology

11.9. protozoology

12. Energy sector
12.1. energy trade

12.2. energy consumption

12.3. energy transmission and distribution

12.4. energy production

12.5. energy storage

12.6. energy policy

12.7. energy technology

13. Pharmacology, pharmacy and pharmaceutical industry
13.1. biopharmacy

13.2. pharmacodynamics

13.3. pharmacogenetics

13.4. pharmacogenomics

13.5. pharmacognosia

13.6. pharmacokinetics

13.7. pharmaceutical chemistry

13.8. pharmaceutical technology

13.9. clinical pharmacology

13.10. drug control

13.11. compounding (<- pharmaceutical compounding; compounding pharmacy)

14. Philosophy
14.1. axiology
14.2. aesthetics
14.3. ethics
14.4. philosophical trends
14.5. philosophy of history
14.6. philosophy of language
14.7. logic
14.8. metaphysics
14.9. philosophy of mind
14.10. legal philosophy
14.11. ontology
14.12. political philosophy
14.13. philosophy of science
14.14. epistemology (<- theory of knowledge)
14.15. philosophy of religion
14.16. social philosophy

15. Folklore studies (<- Folkloristics)

16. Physics
16.1. acoustics
16.2. atomic physics (<- atom physics)
16.3. biophysics
16.4. geophysics
16.5. particle physics
16.6. quantum physics
16.7. mechanics
16.8. molecular physics
16.9. optics
16.10. electrodynamics
16.11. thermodynamics
16.12. nuclear physics

17. Geodesy, cartography and geomatics
17.1. photogrammetry
17.2. geoinformatics
17.3. mapping
17.4. surveying (<- land surveying)
17.5. topography

18. Geology
18.1. geophysics
18.2. geochemistry
18.3. geomorphology
18.4. hydrogeology
18.5. mineralogy and crystallography
18.6. palaeontology
18.7. petrography
18.8. petrology

18.9. sedimentology and stratigraphy
18.10. seismology
18.11. economic geology
18.12. volcanology

19. Administration and management

20. History
20.1. diplomatics
20.2. epigraphy
20.3. ethnohistory
20.4. heraldry
20.5. historical geography
20.6. genealogy
20.7. cultural history
20.8. numismatics
20.9. legal history (<- history of law)
20.10. palaeography (<- paleography)
20.11. papyrology
20.12. political history
20.13. sigillography (<- sphragistics)
20.14. social history
20.15. military history
20.16. economic history
20.17. art history
20.18. history of science
20.19. history of religion

21. Hydrology

22. Engineering sciences
22.1. automation technology
22.2. biotechnology
22.3. chemical engineering
22.4. mechanical engineering
22.5. biomedical engineering
22.6. materials science
22.7. instrument engineering
22.8. nanoscience and nanotechnology
22.9. process engineering
22.10. construction engineering
22.11. electrical engineering
22.12. computer science (<- computing science)
22.13. industrial engineering
22.14. nuclear engineering
22.15. environmental technology

23. Manufacture of footwear

24. Activities of public administration

25. Activities of organizations

25.1. activities of professional membership organizations

25.2. activities of trade unions

25.3. activities of business and employers' membership organizations

25.4. activities of international organizations

25.5. activities of civic organizations

25.6. activities of political organizations

25.7. activities of religious organizations

26. Mining and quarrying

26.1. quarrying of stone, sand and clay

26.2. mining of chemical and fertilizer minerals

26.3. mining of coal and lignite

26.4. mining of metal ores

26.5. extraction of crude petroleum and natural gas

26.6. extraction of salt

26.7. extraction of peat

27. Fishing and fishing industry

27.1. commercial fishing

27.2. fish processing

27.3. fish farming

27.4. crayfishing and crayfish industry

27.5. recreational fishing

28. International relations

28.1. diplomacy

28.2. international trade

28.3. international law

28.4. international policy

28.5. activities of international organizations

28.6. foreign policy

29. Ethnology and ethnography

30. Education and training

30.1. educational science

30.2. education system

30.3. education and training services

30.4. education policy

31. Botany

31.1. economic botany

31.2. plant anatomy

31.3. plant ecology

31.4. plant physiology

31.5. plant genetics

31.6. phytogeography

31.7. plant morphology

31.8. plant systematics

31.9. paleobotany

32. Trade and economy
32.1. customer service
32.2. trading
32.3. accountancy and auditing
32.4. consumption
32.5. marketing
32.6. monetary economy
32.7. finance
32.8. economic history
32.9. economic system
32.10. economic law
32.11. economic policy
32.12. economics
32.13. insurance
32.14. taxation
32.15. rental and leasing activities
32.16. entrepreneurship

33. Chemistry
33.1. analytical chemistry
33.2. biochemistry
33.3. inorganic chemistry
33.4. physical chemistry
33.5. organic chemistry

34. Chemical industry
34.1. chemical engineering
34.2. manufacture of chemicals and chemical products
34.3. manufacture of coke and refined petroleum products
34.4. manufacture of rubber products
34.5. compounding (<- pharmaceutical compounding; compounding pharmacy)
34.6. manufacture of plastic products

35. Linguistics
35.1. dialectology
35.2. ethnolinguistics
35.3. etymology
35.4. phonetics
35.5. phonology
35.6. functional linguistics
35.7. language philosophy
35.8. historical linguistics
35.9. language geography
35.10. grammar
35.11. language policy
35.12. language technology
35.13. language typology
35.14. language teaching
35.15. cognitive linguistics
35.16. translation studies

35.17. lexicology
35.18. lexicography
35.19. neurolinguistics
35.20. legal linguistics
35.21. pragmatics
35.22. psycholinguistics
35.23. semantics
35.24. sociolinguistics
35.25. stylistics
35.26. text linguistics
35.27. terminology
35.28. comparative linguistics

36. Real estate sector
36.1. realty management
36.2. real estate business
36.3. building maintenance and landscape service activities
36.4. real estate agency
36.5. real estate rental and leasing

37. Literary research
37.1. history of literature
37.2. literary criticism

38. Cosmetology and beauty services

39. Household management
39.1. housekeeping
39.2. home economics
39.3. household goods

40. Culture and entertainment
40.1. cultural and entertainment facility operation
40.2. cultural and entertainment event organization
40.3. support activities for culture and entertainment
40.4. cultural policy

41. Visual arts
41.1. graphics
41.2. calligraphy
41.3. sculpture
41.4. painting
41.5. drawing
41.6. art history
41.7. art criticism
41.8. photographic art

42. Applied arts
42.1. ceramic art
42.2. book art
42.3. art of jewellery

42.4. graphic design
42.5. glass art
42.6. metal art
42.7. paper art
42.8. wood art
42.9. textile art

43. Glass and ceramic industry
43.1. manufacture of ceramic products
43.2. manufacture of glass and glass products

44. Transport and logistics
44.1. passenger transport
44.2. freight forwarding
44.3. air transport
44.4. space transport
44.5. cargo handling
44.6. traffic policy
44.7. traffic safety
44.8. moving services
44.9. navigation
44.10. postal and courier services
44.11. pipeline transport (<- transport via pipeline)
44.12. rail transport
44.13. freight transport
44.14. road transport
44.15. customs clearance
44.16. storage
44.17. water transport

45. Physical training and sport
45.1. physical education
45.2. physical training services
45.3. operation of sports facilities and sports club activities
45.4. sport
45.5. sports equipment

46. Medicine and health care (<- Medical sciences and health care)
46.1. anatomy
46.2. anaesthesiology (<- anesthesiology)
46.3. biomedicine
46.4. preventive medicine
46.5. epidemiology
46.6. endocrinology
46.7. physical medicine and rehabilitation (<- physiatry)
46.8. physiology
46.9. gerontology
46.10. gynaecology (<-gynecology)
46.11. dentistry
46.12. haematology (<- hematology)

46.13. hepatology
46.14. methods of treatment
46.15. dermatology
46.16. immunology
46.17. infectious diseases
46.18. oncology
46.19. disaster medicine
46.20. pulmonology
46.21. surgery
46.22. otorhinolaryngology
46.23. paediatrics (<- pediatrics)
46.24. nephrology
46.25. neurology
46.26. legal medicine
46.27. pathology
46.28. medical genetics
46.29. psychiatry
46.30. radiology
46.31. nutritional science
46.32. rheumatology
46.33. sexual medicine
46.34. ophthalmology
46.35. cardiology
46.36. obstetrics
46.37. health care system
46.38. health care services
46.39. health policy
46.40. traumatology
46.41. urology
46.42. alternative medicine
46.43. gastroenterology

47. Geography
47.1. regional geography
47.2. zoogeography
47.3. geomorphology
47.4. historical geography
47.5. plant geography
47.6. cultural geography
47.7. geographic names
47.8. palaeogeography (<- paleogeography)
47.9. political geography
47.10. economic geography
47.11. social geography

48. Soil science

49. Agriculture
49.1. animal husbandry (<- livestock husbandry; livestock management)
49.2. plant cultivation

49.3. agricultural policy
49.4. agricultural science

50. Hotel and catering industry
50.1. food hygiene
50.2. gastronomy
50.3. activities of accommodation providers
50.4. restaurants and other food service activities
50.5. nutritional science
50.6. cuisine

51. Mathematics
51.1. algebra
51.2. mathematical analysis
51.3. arithmetic
51.4. geometry
51.5. number theory
51.6. probability theory
51.7. topology

52. Tourism
52.1. forms of tourism
51.2. travel agency and tour operator activities

53. Metal industry
53.1. machinery industry
53.2. metal products industry

54. Manufacture of basic metals

55. Meteorology and climatology

56. Metrology

57. Hunting and game husbandry

58. Forestry
58.1. multiple-use forestry
58.2. forest ecology
58.3. forest economics
58.4. forest genetics and tree breeding
58.5. forest management (<- silviculture)
58.6. forest protection
58.7. forest research
58.8. forest policy
58.9. forest planning
58.10. forest technology
58.11. logging (<- timber harvesting)
58.12. timber and forest measurement

59. Forest industry
59.1. pulp and paper industry
59.2. wood industry

60. Design

61. Music
61.1. music publishing activities
61.2. music genres
61.3. musicology

62. Leather industry
62.1. manufacture of leather
62.2. manufacture of leather products
62.3. manufacture of leather clothing

63. Law and legislature
63.1. financial law
63.2. administrative law
63.3. international law
63.4. commercial law
63.5. legal services
63.6. legislation
63.7. land law
63.8. rights and freedoms
63.9. legal philosophy
63.10. legal history (<- history of law)
63.11. legal system
63.12. jurisprudence
63.13. family law
63.14. procedural law
63.15. criminal law
63.16. contract law
63.17. labour law (<- employment law)
63.18. constitutional law
63.19. tax law
63.20. water law
63.21. environmental law

64. Genetics
64.1. evolutionary genetics
64.2. genetic engineering
64.3. genomics
64.4. human genetics
64.5. classical genetics
64.6. molecular genetics
64.7. population genetics
64.8. cytogenetics

65. Politics and political research
65.1. international policy
65.2. political philosophy
65.3. political history
65.4. political ideology
65.5. political system

65.6. political economy
65.7. activities of political organizations
65.8. political sociology
65.9. domestic policy
65.10. foreign policy

66. Project management

67. Psychology
67.1. biopsychology
67.2. evolutionary psychology
67.3. psychology of perception
67.4. educational psychology
67.5. developmental psychology
67.6. clinical psychology
67.7. cognitive psychology
67.8. experimental psychology
67.9. cultural psychology
67.10. traffic psychology
67.11. neuropsychology
67.12. legal psychology
67.13. personality psychology
67.14. psycholinguistics
67.15. social psychology
67.16. industrial and organizational psychology (<- I-O psychology; industrial-organizational psychology; work psychology; organizational psychology; work and organizational psychology; industrial psychology; occupational psychology; personnel psychology)
67.17. sport psychology (<- sports psychology)
67.18. psychology of religion
67.19. environmental psychology

68. Garden design and gardening

69. Gambling and betting activities

70. Construction industry and building
70.1. building materials and building products industry
70.2. construction installation
70.3. constructional planning and housing design
70.4. site preparation
70.5. demolition
70.6. building completion and finishing
70.7. architectural engineering (<- building engineering)
70.8. real estate development

71. Semiotics

72. Circus and other entertainment arts
72.1. circus art
72.2. stand up comedy
72.3. impression

73. Warfare and national defence

73.1. weapons and war technology

73.2. military service and non-military service

73.3. armed forces organization and structure

73.4. defence policy

73.5. peacekeeping

73.6. civil defence

73.7. military history

73.8. military law

73.9. art of war

73.10. war economy

73.11. military science

73.12. military administration

73.13. security policy

74. Social security

74.1. social benefits

74.2. social welfare

74.3. social services

74.4. social policy

74.5. social insurance

75. Sociology

75.1. historical sociology

75.2. sociology of education (<- educational sociology)

75.3. urban sociology

75.4. sociology of culture

75.5. sociology of consumption

75.6. rural sociology

75.7. sociology of law

75.8. sociology of the family

75.9. political sociology

75.10. social history

75.11. sociolinguistics

75.12. military sociology

75.13. sociology of art

75.14. economic sociology

75.15. sociology of health and illness

75.16. sociology of knowledge

75.17. sociology of science

75.18. sociology of work (<- industrial sociology)

75.19. sociology of communication

75.20. sociology of religion

75.21. environmental sociology

76. Standardization

77. Electrical and electronics industry

77.1. manufacture of electronic components

77.2. manufacture of electromedical and electrotherapeutic equipment

77.3. manufacture of appliances for measuring and testing
77.4. manufacture of appliances for navigation
77.5. manufacture of optical instruments
77.6. manufacture of electrical equipment
77.7. manufacture of irradiation equipment
77.8. manufacture of data carriers
77.9. manufacture of computers and peripheral equipment
77.10. manufacture of photographic equipment
77.11. manufacture of communication equipment
77.12. manufacture of consumer electronics

78. Dance
78.1. choreography
78.2. dance styles (<- styles of dance)

79. Theatre and film sector (<- Theatre and cinema sector)
79.1. scenic design (<- scenography)
79.2. make-up
79.3. acting technique
79.4. direction
79.5. properties (<- theatrical properties)
79.6. theatre and film criticism
79.7. theatrical genres
79.8. theatre research
79.9. lighting design
79.10. video and film production
79.11. sound design

80. Technical services
80.1. architectural services
80.2. installation, repair and maintenance services
80.3. engineering services
80.4. technical consultancy
80.5. technical inspection
80.6. technical testing and analysis

81. Textile industry
81.1. weaving of textiles
81.2. finishing of textiles
81.3. preparation and spinning of textile fibres
81.4. manufacture of textile products
81.5. manufacture of textile clothing

82. Study of science (<- Science studies)
82.1. ethics of science
82.2. philosophy of science
82.3. history of science
82.4. sociology of science

83. Information compilation and storage
83.1. archive-keeping

83.2. archival science
83.3. library services
83.4. library science
83.5. museum activities
83.6. museology (<- museum studies)

84. Data processing and information technology
84.1. computer science (<- computing science)
84.2. information and communication technology (<- ICT)
84.3. information technology services

85. Statistics

86. Toxicology

87. Tobacco industry

88. Fur industry
88.1. dressing and dyeing of fur
88.2. manufacture of articles of fur
88.3. manufacture of fur clothing

89. Safety and security
89.1. investigation activities
89.2. close protection
89.3. emergency centre activities
89.4. maintenance of order
89.5. traffic safety (<- road safety)
89.6. chimney sweeping
89.7. fire services
89.8. rescue services
89.9. police
89.10. border control
89.11. risk management
89.12. radiation protection
89.13. intelligence services (<- intelligence agencies)
89.14. security check services
89.15. safety and security technology
89.16. occupational safety (<- industrial safety; safety at work)
89.17. correctional services

90. Research and development
90.1. research and development infrastructure
90.2. research policy
90.3. research support activities

91. Working life and labour protection
91.1. career counseling (<- career guidance)
91.2. activities of professional membership organizations
91.3. activities of trade unions
91.4. activities of business and employers' membership organizations
91.5. ergonomics

91.6. human resource management (<- HMR)
91.7. training of personnel (<- personnel training)
91.8. provision of human resources
91.9. employment activities
91.10. work organization
91.11. labour law (<- employment law)
91.12. occupational health care
91.13. occupational safety (<- industrial safety; safety at work)
91.14. labour policy

92. Astronomy and space science
92.1. astrophysics
92.2. astrometry
92.3. cosmology
92.4. celestial mechanics

93. Beliefs and occultism
93.1. alchemy
93.2. anthroposophy
93.3. astrology
93.4. fortune-telling
93.5. fengshui
93.6. Kabbalah
93.7. magic
93.8. numerology
93.9. parapsychology
93.10. spiritualism
93.11. theosophy
93.12. ufology

94. Religion
94.1. activities of religious organizations
94.2. theology
94.3. philosophy of religion
94.4. religions and religious trends
94.5. history of religion
94.6. religious education
94.7. religious studies

95. Leisure and hobbies

96. Water supply, sewerage and waste management
96.1. waste collection
96.2. waste treatment and disposal
96.3. materials recovery
96.4. water collection, treatment and supply
96.5. sewerage

97. Communication and dissemination of information
97.1. mass media activities
97.2. journalism

97.3. publishing activities

97.4. printing and reproduction of recorded media

97.5. radio and television broadcasting and programming activities

97.6. telecommunication

97.7. information services

97.8. communication policy

97.9. communication sciences

98. Demography

Appendix 3.3   Domain classification in Russian

1. Общие термины

2. Неклассифицированные тематические области
2.1. неклассифицированная область знаний
2.2. неклассифицированная область деятельности

3. Антропология

4. Археология

5. Архитектура
5.1. история архитектуры
5.2. теория архитектуры
5.3. ландшафтная архитектура
5.4. проектирование конструкций
5.5. проектирование зданий и жилищное проектирование
5.6. дизайн интерьера
5.7. территориальное планирование

6. Биология
6.1. анатомия
6.2. биофизика
6.3. биохимия
6.4. биологическая антропология (<- физическая антропология)
6.5. биотехнология
6.6. зоология
6.7. этология
6.8. эволюционная биология
6.9. физиология
6.10. гистология
6.11. ботаника
6.12. биология развития
6.13. микробиология
6.14. молекулярная биология
6.15. палеонтология
6.16. генетика
6.17. микология
6.18. цитология (<- клеточная биология; биология клетки)
6.19. биология окружающей среды

7. Экология и охрана окружающей среды

8. Производство продуктов питания
8.1. производство готовых пищевых продуктов
8.2. пищевые технологии
8.3. производство готовых кормов для животных

8.4. переработка и консервирование фруктов, ягод и овощей

8.5. производство гомогенизированных продуктов и диетических пищевых продуктов

8.6. производство напитков

8.7. переработка и консервирование рыбы, ракообразных и моллюсков

8.8. производство какао, шоколада и сахаристых кондитерских изделий

8.9. производство растительных и животных масел и жиров

8.10. производство мучных и хлебобулочных изделий

8.11. производство продуктов из мяса и мяса домашней птицы

8.12. производство молочных продуктов

8.13. производство пряностей и приправ

8.14. производство продуктов мукомольной промышленности, крахмалов и крахмальных продуктов

8.15. производство сахара

8.16. переработка чая и кофе

8.17. переработка и консервирование мяса

9. Ветеринария и ветеринарные услуги

10. Уход за животными и защита животных

11. Зоология
11.1. арахнология

11.2. зоогеография

11.3. герпетология

11.4. энтомология

11.5. ихтиология

11.6. орнитология

11.7. маммалогия

11.8. палеозоология

11.9. протозоология

12. Энергетика
12.1. торговля в области энергетики

12.2. использование энергии

12.3. передача и распределение энергии

12.4. производство энергии

12.5. хранение энергии

12.6. энергетическая политика

12.7. энергетические технологии

13. Фармакология, фармацевтика и фармацевтическая промышленность (<- Фармакология, фармация и фармацевтическая промышленность)
13.1. биофармацевтика (<- биофармация)

13.2. фармакодинамика

13.3. фармакогенетика

13.4. фармакогеномика

13.5. фармакогнозия

13.6. фармакокинетика

13.7. фармацевтическая химия

13.8. фармацевтическая технология

13.9. клиническая фармакология
13.10. контроль качества лекарственных средств
13.11. производство лекарственных средств

14. Философия
14.1. аксиология
14.2. эстетика
14.3. этика
14.4. философские течения
14.5. философия истории
14.6. философия языка
14.7. логика
14.8. метафизика
14.9. философия сознания
14.10. философия права
14.11. онтология
14.12. политическая философия
14.13. философия науки
14.14. эпистемология (<- гносеология; теория познания)
14.15. философия религии
14.16. социальная философия

15. Фольклористика

16. Физика
16.1. акустика
16.2. атомная физика
16.3. биофизика
16.4. геофизика
16.5. физика элементарных частиц
16.6. квантовая физика
16.7. механика
16.8. молекулярная физика
16.9. оптика
16.10. электродинамика
16.11. термодинамика
16.12. ядерная физика

17. Геодезия, картография и геоматика
17.1. фотограмметрия
17.2. геоинформатика
17.3. картографирование
17.4. геодезические работы
17.5. топография

18. Геология
18.1. геофизика
18.2. геохимия
18.3. геоморфология
18.4. гидрогеология
18.5. минералогия и кристаллография

18.6. палеонтология

18.7. петрография

18.8. петрология

18.9. седиментология и стратиграфия

18.10. сейсмология

18.11. экономическая геология

18.12. вулканология

19. Администрирование и менеджмент

20. История

20.1. дипломатика

20.2. эпиграфика

20.3. этническая история

20.4. геральдика

20.5. историческая география

20.6. генеалогия

20.7. история культуры

20.8. нумизматика

20.9. история права

20.10. палеография

20.11. папирология

20.12. политическая история

20.13. сфрагистика (<- сигиллография)

20.14. социальная история

20.15. военная история

20.16. история экономики

20.17. история искусств

20.18. история науки

20.19. история религии

21. Гидрология

22. Технические науки

22.1. технологии автоматизации

22.2. биотехнология

22.3. химическая технология

22.4. инженерная механика

22.5. биомедицинская инженерия

22.6. материаловедение

22.7. приборостроение

22.8. нанонаука и нанотехнологии

22.9. технология производства

22.10. строительный инжиниринг

22.11. электротехника

22.12. информатика

22.13. промышленная инженерия (<- производственная инженерия)

22.14. ядерная техника

22.15. экологический инжиниринг

23. Производство обуви (<- Обувная промышленность)

24. Государственное и местное управление

25. Деятельность организаций
25.1. деятельность профессиональных членских организаций
25.2. деятельность профсоюзов
25.3. деятельность членских организаций предпринимателей и объединений работодателей
25.4. деятельность международных организаций
25.5. деятельность гражданских организаций
25.6. деятельность политических организаций
25.7. деятельность религиозных организаций

26. Горнодобывающая промышленность и разработка карьеров
26.1. разработка каменных, глиняных и песчаных карьеров
26.2. добыча минерального сырья для химической промышленности и производства удобрений
26.3. добыча угля и лигнита
26.4. добыча металлических руд
26.5. добыча сырой нефти и природного газа
26.6. добыча соли
26.7. добыча торфа

27. Рыбалка, рыболовство и рыбное хозяйство
27.1. рыболовство
27.2. переработка рыбы
27.3. рыбоводство
27.4. ловля и разведение раков
27.5. рыбалка

28. Международные отношения
28.1. дипломатия
28.2. международная торговля
28.3. международное право
28.4. международная политика
28.5. деятельность международных организаций
28.6. внешняя политика

29. Этнология и этнография

30. Образование и обучение
30.1. педагогика
30.2. система образования
30.3. услуги образования и обучения
30.4. образовательная политика

31. Ботаника
31.1. ботаническое ресурсоведение (<- экономическая ботаника)
31.2. анатомия растений (<- гистология растений)
31.3. экология растений
31.4. физиология растений
31.5. генетика растений
31.6. география растений (<- фитогеография)
31.7. морфология растений

31.8. систематика растений

31.9. палеоботаника

32. Торговля и экономика

32.1. обслуживание клиентов

32.2. торговля

32.3. бухгалтерский учет и аудит

32.4. потребление

32.5. маркетинг

32.6. финансы

32.7. финансово-инвестиционная деятельность

32.8. история экономики

32.9. экономическая система

32.10. экономическое право

32.11. экономическая политика

32.12. экономика (<- экономическая наука)

32.13. страхование

32.14. налогообложение

32.15. аренда и лизинг

32.16. предпринимательская деятельность

33. Химия

33.1. аналитическая химия

33.2. биохимия

33.3. неорганическая химия

33.4. физическая химия

33.5. органическая химия

34. Химическая промышленность

34.1. химическая технология

34.2. производство химикатов и химической продукции

34.3. производство кокса и продуктов нефтепереработки

34.4. производство резиновых изделий

34.5. производство лекарственных средств

34.6. производство пластмассовых изделий

35. Лингвистика

35.1. диалектология

35.2. этнолингвистика

35.3. этимология

35.4. фонетика

35.5. фонология

35.6. функциональная лингвистика

35.7. философия языка

35.8. историческая лингвистика

35.9. лингвистическая география (<- лингвогеография)

35.10. грамматика

35.11. языковая политика

35.12. лингвистические технологии

35.13. типология языков (<- лингвистическая типология)

35.14. обучение языкам
35.15. когнитивная лингвистика
35.16. переводоведение
35.17. лексикология
35.18. лексикография
35.19. нейролингвистика
35.20. юридическая лингвистика
35.21. прагматика
35.22. психолингвистика
35.23. семантика
35.24. социолингвистика
35.25. стилистика
35.26. лингвистика текста
35.27. терминология
35.28. сравнительная лингвистика

36. Недвижимость
36.1. управление недвижимостью
36.2. торговля недвижимостью
36.3. обслуживание зданий и территорий
36.4. риэлторская деятельность
36.5. аренда недвижимости

37. Литературоведение
37.1. история литературы
37.2. литературная критика

38. Косметология и косметологические услуги

39. Домашнее хозяйство
39.1. ведение домашнего хозяйства
39.2. домоводство
39.3. хозяйственные товары

40. Культура и развлечения
40.1. деятельность культурных и развлекательных учреждений
40.2. организация культурных и развлекательных мероприятий
40.3. техническая поддержка в области культуры и развлечений
40.4. культурная политика

41. Изобразительное искусство
41.1. графика
41.2. каллиграфия
41.3. скульптура
41.4. живопись
41.5. рисунок (<- академический рисунок)
41.6. история искусств
41.7. художественная критика (<- арт-критика)
41.8. искусство фотографии (<- фотоискусство)

42. Декоративно-прикладное искусство

42.1. искусство керамики

42.2. книжное искусство

42.3. ювелирное искусство

42.4. графический дизайн

42.5. художественное стеклоделие

42.6. художественная обработка металла

42.7. бумажное искусство

42.8. художественная обработка дерева

42.9. художественная обработка текстиля (<- текстильное искусство)

43. Стекольно-керамическая промышленность

43.1. производство керамических изделий

43.2. производство стекла и изделий из стекла

44. Транспорт и логистика

44.1. пассажирские перевозки

44.2. экспедирование грузов

44.3. воздушные перевозки

44.4. космический транспорт

44.5. обработка грузов

44.6. транспортная политика

44.7. транспортная безопасность

44.8. услуги по переезду (<- услуги переезда)

44.9. навигация

44.10. почтовые и курьерские услуги

44.11. трубопроводная транспортировка

44.12. железнодорожные перевозки

44.13. грузовые перевозки

44.14. автотранспортные перевозки

44.15. таможенная очистка грузов

44.16. хранение грузов

44.17. водные перевозки

45. Физкультура и спорт

45.1. физвоспитание

45.2. услуги физкультуры и спорта

45.3. деятельность спортивных объектов и спортивных клубов

45.4. виды спорта

45.5. спортивное оборудование

46. Медицина и здравоохранение

46.1. анатомия

46.2. анестезиология

46.3. биомедицина

46.4. профилактическая медицина

46.5. эпидемиология

46.6. эндокринология

46.7. физиатрия

46.8. физиология

46.9. геронтология

46.10. гинекология

46.11. стоматология

46.12. гематология

46.13. гепатология

46.14. методы лечения

46.15. дерматология

46.16. иммунология

46.17. инфекционные заболевания

46.18. онкология

46.19. медицина катастроф

46.20. пульмонология

46.21. хирургия

46.22. оториноларингология (<- отоларингология)

46.23. педиатрия

46.24. нефрология

46.25. неврология

46.26. судебная медицина

46.27. патология

46.28. медицинская генетика

46.29. психиатрия

46.30. радиология

46.31. нутрициология

46.32. ревматология

46.33. сексуальная медицина

46.34. офтальмология

46.35. кардиология

46.36. акушерство

46.37. система здравоохранения

46.38. услуги здравоохранения

46.39. политика в сфере здравоохранения

46.40. травматология

46.41. урология

46.42. альтернативная медицина

46.43. гастроэнтерология

47. География

47.1. страноведение

47.2. зоогеография

47.3. геоморфология

47.4. историческая география

47.5. география растений (<- фитогеография)

47.6. культурная география

47.7. географические названия

47.8. палеогеография

47.9. политическая география

47.10. экономическая география

47.11. социальная география

48. Почвоведение

49. Сельское хозяйство
49.1. животноводство
49.2. растениеводство
49.3. сельскохозяйственная политика
49.4. агрономия

50. Гостиничный бизнес и кейтеринг
50.1. гигиена питания
50.2. гастрономия
50.3. деятельность учреждений, предоставляющих услуги проживания
50.4. деятельность учреждений, предоставляющих услуги питания
50.5. диетология
50.6. кулинарное искусство

51. Математика
51.1. алгебра
51.2. математический анализ (<- матанализ)
51.3. арифметика
51.4. геометрия
51.5. теория чисел
51.6. теория вероятности
51.7. топология

52. Туризм
52.1. виды туризма
52.2. деятельность туроператоров и турагентов

53. Металлообрабатывающая промышленность
53.1. машиностроение
53.2. производство металлопродукции

54. Металлургическая промышленность (<- Металлургия)

55. Метеорология и климатология

56. Метрология

57. Охота и охотничье хозяйство

58. Лесное хозяйство
58.1. многоцелевое использование лесов
58.2. лесная экология
58.3. экономика лесопользования
58.4. лесная генетика и селекция
58.5. лесовыращивание
58.6. защита леса (<- лесозащита)
58.7. исследования леса
58.8. лесная политика
58.9. лесохозяйственное планирование
58.10. технология лесохозяйственных работ

58.11. лесосечные работы

58.12. обмер деревьев и леса на корню

59. Лесная промышленность

59.1. целлюлозно-бумажная промышленность

59.2. деревообрабатывающая промышленность

60. Дизайн

61. Музыка

61.1. деятельность по изданию фонограмм и музыкальных записей

61.2. музыкальные жанры (<- музыкальные стили)

61.3. музыковедение

62. Кожевенная промышленность

62.1. производство кожи

62.2. производство изделий из кожи

62.3. пошив кожаной одежды

63. Право и законодательство

63.1. финансовое право

63.2. административное право

63.3. международное право

63.4. торговое право (<- коммерческое право)

63.5. юридические услуги

63.6. законодательная деятельность

63.7. земельное право

63.8. права и свободы

63.9. философия права

63.10. история права

63.11. правовая система

63.12. юриспруденция (<- правоведение)

63.13. семейное право

63.14. процессуальное право

63.15. уголовное право

63.16. договорное право

63.17. трудовое право

63.18. конституционное право

63.19. налоговое право

63.20. водное право

63.21. экологическое право

64. Генетика

64.1. эволюционная генетика

64.2. генетическая инженерия (<- генная инженерия)

64.3. геномика

64.4. генетика человека

64.5. классическая генетика

64.6. молекулярная генетика

64.7. популяционная генетика

64.8. цитогенетика

65. Политика и политология

65.1. международная политика

65.2. политическая философия

65.3. политическая история

65.4. политическая идеология

65.5. политическая система

65.6. политическая экономика

65.7. деятельность политических организаций

65.8. политическая социология

65.9. внутренняя политика

65.10. внешняя политика

66. Управление проектами

67. Психология

67.1. биопсихология

67.2. эволюционная психология

67.3. психология восприятия

67.4. педагогическая психология (<- психология образования)

67.5. психология развития (<- возрастная психология)

67.6. клиническая психология

67.7. когнитивная психология

67.8. экспериментальная психология

67.9. культурная психология

67.10. транспортная психология

67.11. нейропсихология

67.12. юридическая психология (<- правовая психология)

67.13. психология личности

67.14. психолингвистика

67.15. социальная психология

67.16. организационная психология (<- психология менеджмента)

67.17. психология спорта (<- спортивная психология)

67.18. психология религии

67.19. психология среды

68. Ландшафтный дизайн и садоводство

69. Азартные игры и заключение пари

70. Строительная промышленность и строительство

70.1. производство строительных материалов и изделий

70.2. строительно-монтажные работы

70.3. проектирование зданий и жилищное проектирование

70.4. подготовительные работы на стройплощадке

70.5. снос зданий и сооружений

70.6. отделочные строительные работы

70.7. технологии строительства (<- строительные технологии)

70.8. застройка

71. Семиотика

72. Цирк и другие виды эстрадного искусства

72.1. цирковое искусство

72.2. юмористическая эстрада (<- стендап камеди)

72.3. пародия

73. Военные действия и национальная оборона

73.1. оружие и военная техника

73.2. военная и гражданская служба

73.3. организация и структура вооруженных сил

73.4. оборонная политика (<- военная политика)

73.5. миротворческая деятельность

73.6. гражданская оборона

73.7. военная история

73.8. военное право

73.9. военное искусство

73.10. военная экономика

73.11. военная наука

73.12. военная администрация

73.13. политика в области безопасности

74. Социальное обеспечение

74.1. социальные льготы

74.2. социальное обслуживание

74.3. социальные услуги

74.4. социальная политика

74.5. социальное страхование

75. Социология

75.1. историческая социология

75.2. социология образования

75.3. социология города

75.4. социология культуры

75.5. социология потребления

75.6. социология села

75.7. социология права

75.8. социология семьи

75.9. политическая социология

75.10. социальная история

75.11. социолингвистика

75.12. военная социология

75.13. социология искусства

75.14. экономическая социология

75.15. социология здоровья

75.16. социология знания

75.17. социология науки

75.18. социология труда

75.19. социология коммуникации

75.20. социология религии

75.21. экологическая социология (<- экосоциология)

76. Стандартизация

77. Электротехническая и электронная промышленность
77.1. производство электронных деталей
77.2. производство электромедицинского и электротерапевтического оборудования
77.3. производство приборов для измерения и тестирования
77.4. производство навигационных приборов и оборудования
77.5. производство оптических приборов
77.6. производство электрического оборудования
77.7. производство оборудования для облучения
77.8. производство носителей информации
77.9. производство компьютеров и периферийного оборудования
77.10. производство фотографического оборудования
77.11. производство коммуникационного оборудования
77.12. производство бытовой электроники

78. Танец
78.1. хореография
78.2. направления и стили танца

79. Театр и кино
79.1. сценография
79.2. грим (<- гримировальное искусство)
79.3. актерское искусство
79.4. режиссура
79.5. реквизит
79.6. театральная и кинокритика
79.7. жанры театрального искусства (<- театральные жанры)
79.8. театроведение
79.9. световая сценография
79.10. производство видеофильмов и кинопромышленность
79.11. звукорежиссура

80. Технические услуги
80.1. архитектурные услуги
80.2. услуги по установке, ремонту и техническому обслуживанию
80.3. инжиниринг
80.4. технические консультации
80.5. технические проверки
80.6. технические испытания и анализы

81. Текстильная промышленность
81.1. изготовление тканей
81.2. отделка текстильных изделий
81.3. подготовка и прядение текстильных волокон
81.4. производство текстильных изделий
81.5. производство одежды из текстиля

82. Науковедение
82.1. этика науки
82.2. философия науки

82.3. история науки

82.4. социология науки

83. Сбор и хранение информации

83.1. архивное дело

83.2. архивоведение

83.3. библиотечные услуги

83.4. библиотековедение

83.5. музейная деятельность (<- деятельность музеев)

83.6. музееведение (<- музеология)

84. Обработка данных и информационные технологии

84.1. информатика

84.2. информационно-коммуникационные технологии (<- ИКТ)

84.3. услуги по обработке данных

85. Статистика

86. Токсикология

87. Табачная промышленность

88. Меховая промышленность

88.1. выделка и окрашивание меха

88.2. производство меховых изделий (<- производство изделий из меха)

88.3. производство одежды из меха (<- производство меховой одежды)

89. Обеспечение безопасности

89.1. детективная деятельность

89.2. обеспечение личной безопасности

89.3. деятельность экстренных центров

89.4. поддержание общественного порядка

89.5. транспортная безопасность

89.6. чистка дымоходов

89.7. пожарная охрана

89.8. деятельность службы спасения

89.9. деятельность полиции

89.10. пограничный контроль

89.11. управление рисками

89.12. радиационная защита

89.13. деятельность спецслужб

89.14. контроль безопасности

89.15. технологии безопасности

89.16. безопасность труда

89.17. деятельность исправительных учреждений

90. Научные исследования и развитие

90.1. научно-исследовательская инфраструктура

90.2. политика в области науки

90.3. техническая поддержка в области науки

91. Трудовые отношения и охрана труда

91.1. профессиональная ориентация (<- профориентация)

91.2. деятельность профессиональных членских организаций

91.3. деятельность профсоюзов

91.4. деятельность членских организаций предпринимателей и объединений работодателей

91.5. эргономика

91.6. управление персоналом

91.7. обучение персонала

91.8. обеспечение персоналом

91.9. трудоустройство

91.10. организация труда

91.11. трудовое право

91.12. трудовое здравоохранение

91.13. безопасность труда

91.14. политика занятости

92. Астрономия и исследования космоса

92.1. астрофизика

92.2. астрометрия

92.3. космология

92.4. небесная механика

93. Верования и оккультизм

93.1. алхимия

93.2. антропософия

93.3. астрология

93.4. предсказания

93.5. фен-шуй

93.6. Каббала

93.7. магия

93.8. нумерология

93.9. парапсихология

93.10. спиритуализм

93.11. теософия

93.12. уфология

94. Религия

94.1. деятельность религиозных организаций

94.2. теология

94.3. философия религии

94.4. мировые религии и религиозные течения

94.5. история религии

94.6. религиозное образование

94.7. религиоведение

95. Досуг и хобби

96. Водоснабжение, канализация и управление отходами

96.1. сбор отходов

96.2. обработка и удаление отходов

96.3. вторичная переработка

96.4. забор, подготовка и распределение воды

96.5. канализация и очистка сточных вод

97. Коммуникации и передача информации

97.1. деятельность средств массовой информации (<- деятельность СМИ)

97.2. журналистика

97.3. издательская деятельность

97.4. полиграфия и тиражирование записанных носителей информации

97.5. производство программ и телерадиовещание

97.6. телекоммуникации

97.7. информационные услуги

97.8. политика в области коммуникаций

97.9. коммуникативные науки (<- коммуникационная наука; теория коммуникации)

98. Демография

1. Allgemeinbegriffe

2. Nicht klassifizierte Domänen
2.1. nicht klassifizierter Fachbereich
2.2. nicht klassifizierter Tätigkeitsbereich

3. Anthropologie

4. Archäologie

5. Architektur
5.1. Architekturgeschichte
5.2. Architekturtheorie
5.3. Landschaftsarchitektur
5.4. Bautechnik
5.5. Gebäude- und Wohnungsplanung
5.6. Raumgestaltung
5.7. Regionalplanung

6. Biologie
6.1. Anatomie
6.2. Biophysik
6.3. Biochemie
6.4. Biologische Anthropologie (<- Physische Anthropologie)
6.5. Biotechnologie
6.6. Tierkunde (<- Zoologie)
6.7. Ethologie (<- Verhaltensforschung)
6.8. Evolutionsbiologie
6.9. Physiologie
6.10. Histologie (<- Gewebelehre)
6.11. Pflanzenkunde (<- Botanik)
6.12. Entwicklungsbiologie
6.13. Mikrobiologie
6.14. Molekularbiologie
6.15. Paläontologie
6.16. Vererbungslehre (<- Genetik)
6.17. Pilzkunde
6.18. Zellbiologie
6.19. Umweltbiologie

7. Ökologie und Umweltschutz

8. Ernährungsindustrie
8.1. Herstellung von Halbfertig- und Fertiggerichten
8.2. Ernährungstechnologie
8.3. Herstellung von Futtermitteln
8.4. Obst-, Beeren- und Gemüseverarbeitung

8.5. Herstellung von homogenisierten und diätetischen Nahrungsmitteln

8.6. Getränkeherstellung

8.7. Fischverarbeitung

8.8. Herstellung von Kakao, Schokolade und Süßwaren

8.9. Herstellung von pflanzlichen und tierischen Ölen und Fetten

8.10. Herstellung von Back- und Teigwaren

8.11. Fleisch- und Geflügelverarbeitung

8.12. Milchverarbeitung

8.13. Herstellung von Würzmitteln und Soßen

8.14. Herstellung von Mehl und Stärkeerzeugnissen

8.15. Herstellung von Zucker

8.16. Verarbeitung von Kaffee und Tee

8.17. Schlachtung und Fleischverarbeitung

9. Tiermedizin und Veterinärleistungen

10. Tierpflege und Tierschutz

11. Tierkunde (<- Zoologie)
11.1. Arachnologie (<- Spinnenforschung)

11.2. Tiergeographie

11.3. Herpetologie

11.4. Insektenkunde

11.5. Fischkunde

11.6. Vogelkunde

11.7. Mammologie (<- Säugetierkunde)

11.8. Paläozoologie

11.9. Protozoologie

12. Energiebereich
12.1. Energiehandel

12.2. Energieversorgung

12.3. Energieübertragung und Verteilung

12.4. Energieerzeugung

12.5. Energiespeicherung

12.6. Energiepolitik

12.7 Energietechnologie

13. Pharmakologie, Pharmazie und Pharmaindustrie
13.1. Biopharmazie

13.2. Pharmakodynamik

13.3. Pharmakogenetik

13.4. Pharmakogenomik

13.5. Pharmakognosie

13.6. Pharmakokinetik

13.7. pharmazeutische Chemie

13.8. pharmazeutische Technologie

13.9. Klinische Pharmakologie

13.10. Arzneimittelkontrolle

13.11. Arzneimittelherstellung

14. Philosophie
14.1. Wertetheorie (<- Wertelehre; Wertephilosophie; Axiologie)
14.2. Ästhetik
14.3. Ethik
14.4. philosophische Trends
14.5. Geschichtsphilosophie
14.6. Sprachphilosophie
14.7. Logik
14.8. Metaphysik
14.9. Philosophie des Verstandes
14.10. Rechtsphilosophie
14.11. Ontologie
14.12. Politische Philosophie
14.13. Wissenschaftsphilosophie
14.14. Erkenntnislehre (<- Wissenschaftslehre)
14.15. Religionsphilosophie
14.16. Sozialphilosophie

15. Volkskunde

16. Physik
16.1. Akustik
16.2. Atomphysik
16.3. Biophysik
16.4. Geophysik
16.5. Teilchenphysik
16.6. Quantenphysik
16.7. Mechanik
16.8. Molekularphysik
16.9. Optik
16.10. Elektrodynamik
16.11. Thermodynamik
16.12. Kernphysik

17. Geodesie, Kartographie und Geomatik (<- Vermessungskunde, Kartenkunde und Geomatik)
17.1. Messbildverfahren
17.2. Geoinformationssystem
17.3. Kartierung
17.4. Landvermessung
17.5. Geländekunde

18. Geologie
18.1. Geophysik
18.2. Geochemie
18.3. Geomorphologie
18.4. Hydrogeologie
18.5. Mineralogie und Kristallographie
18.6. Paläontologie
18.7. Petrographie
18.8. Petrologie

25. Organisations- und Vereinstätigkeit

25.1. Berufsorganisationen

25.2. Gewerkschaften

25.3. Wirtschafts- und Arbeitgeberverbände

25.4. internationale Organisationen

25.5. Bürgervereine

25.6. politische Organisationen

25.7. kirchliche und sonstige religiöse Vereinigungen

26. Bergbau

26.1. Förderung von Naturstein, Kies, Sand, Ton

26.2. Förderung von chemischen und Düngemittelmineralen

26.3. Stein- und Braunkohlebergbau

26.4. Erzbergbau

26.5. Gewinnung von Erdöl und Erdgas

26.6. Gewinnung von Salz

26.7. Torfgewinnung

27. Fischerei und Fischindustrie

27.1. gewerbliche Fischerei

27.2. Fischverarbeitung

27.3. Fischzucht

27.4. Krebswirtschaft

27.5. Freizeitfischerei

28. Auslandbeziehungen

28.1. Diplomatie

28.2. internationaler Handel

28.3. Internationales Recht

28.4. internationale Politik

28.5. internationale Organisationen

28.6. Außenpolitik

29. Ethnologie und Ethnografie

30. Erziehung und Bildung

30.1. Erziehungswissenschaft

30.2. Bildungssystem

30.3. Aus- und Fortbildung

30.4. Bildungspolitik

31. Pflanzenkunde (<- Botanik)

31.1. Ökonomische Pflanzenkunde

31.2. Pflanzenanatomie

31.3. Pflanzenökologie

31.4. Pflanzenphysiologie

31.5. Pflanzengenetik

31.6. Pflanzengeographie (<- Geobotanik)

31.7. Pflanzenmorphologie

31.8. Pflanzensystematik

31.9. Paläobotanik

32. Wirtschaft und Handel

32.1. Kundenservice

32.2. Handelsverkehr

32.3. Buchführung und Rechnungswesen

32.4. Konsum

32.5. Marketing

32.6. Geldwirtschaft

32.7. Finanzwirtschaft

32.8. Wirtschaftsgeschichte

32.9. Wirtschaftssystem

32.10. Wirtschaftsrecht

32.11. Wirtschaftspolitik

32.12. Wirtschaftswissenschaft

32.13. Versicherungswesen

32.14. Steuerwesen

32.15. Miet- und Leasingaktivitäten

32.16. Geschäftsführung

33. Chemie

33.1. Analytische Chemie

33.2. Biochemie

33.3. Anorganische Chemie

33.4. Physikalische Chemie

33.5. Organische Chemie

34. Chemieindustrie

34.1. Chemieingenieurswesen

34.2. Herstellung von chemischen Erzeugnissen

34.3. Kokerei und Mineralölverarbeitung

34.4. Herstellung von Gummiwaren

34.5. Arzneimittelherstellung

34.6. Herstellung von Kunststoffwaren

35. Sprachwissenschaft (<- Linguistik)

35.1. Dialektologie

35.2. Ethnolinguistik

35.3. Etymologie (<- Wortherkunft)

35.4. Fonetik

35.5. Fonologie

35.6. Funktionssprachwissenschaft

35.7. Sprachphilosophie

35.8. Historische Linguistik (<- Sprachgeschichte)

35.9. Sprachgeographie

35.10. Grammatik

35.11. Sprachpolitik

35.12. Sprachtechnologie

35.13. Sprachtypologie

35.14. Sprachunterricht

35.15. Kognitive Linguistik

35.16. Translationswissenschaft

60. Gestaltung

61. Musik
61.1. Musikverlagstätigkeiten
61.2. Musikgattungen
61.3. Musikwissenschaft

62. Lederindustrie
62.1. Lederherstellung
62.2. Herstellung von Lederprodukten
62.3. Herstellung von Lederkleidung

63. Recht und Legislative
63.1. Finanzrecht
63.2. Verwaltungsrecht
63.3. Internationales Recht
63.4. Handelsrecht
63.5. Rechtsdienstleistungen
63.6. Gesetzgebung
63.7. Bodenrecht
63.8. Rechte und Freiheiten
63.9. Rechtsphilosophie
63.10. Rechtsgeschichte (<- Justizgeschichte)
63.11. Rechtsordnung
63.12. Rechtslehre
63.13. Familienrecht
63.14. Prozessrecht
63.15. Strafrecht
63.16. Vertragsrecht
63.17. Arbeitsrecht
63.18. Staatsrecht
63.19. Steuerrecht
63.20. Wasserrecht
63.21. Umweltrecht

64. Genetik
64.1. Evolutionsgenetik
64.2. Gentechnologie
64.3. Genomforschung
64.4. Humangenetik
64.5. Klassische Genetik
64.6. Molekulargenetik
64.7. Populationsgenetik
64.8. Cytogenetik

65. Politik und Politikforschung
65.1. Internationale Politik
65.2. Politische Philosophie
65.3. Politikgeschichte (<- Politische Geschichte)
65.4. politische Ideologie
65.5. politische Ordnung

77.5. Herstellung von optischen Instrumenten und Geräten

77.6. Herstellung von Elektrogeräten

77.7. Herstellung von Bestrahlungsgeräten

77.8. Herstellung von Datenträgern

77.9. Herstellung von Datenverarbeitungsgeräten und peripheren Geräten

77.10. Herstellung von fotografischen Instrumenten und Geräten

77.11. Herstellung von Geräten und Einrichtungen der Telekommunikation

77.12. Herstellung von Geräten der Unterhaltungselektronik

78. Tanzkunst

78.1. Choreographie

78.2. Tanzstile (<- Tanzweise)

79. Theater- und Filmsektor (<- Theater- und Kinobereich)

79.1. Bühnenbild (<- Szenographie)

79.2. Maskenbild

79.3. Schauspielkunst

79.4. Regie

79.5. Requisiten

79.6. Theater- und Filmkritik

79.7. schauspielerische Gattungen

79.8. Theaterwissenschaft

79.9. Lichtgestaltung

79.10. Video- und Filmproduktion

79.11. Tongestaltung

80. Technische Dienstleistungen

80.1. Architekturservice

80.2. Installations-, Reparatur- und Wartungsdienstleistungen

80.3. Ingenieurdienstleistungen

80.4. technische Beratung

80.5. technische Prüfung

80.6. technisches Testen und technische Analyse

81. Textilindustrie

81.1. Webekunst

81.2. Veredelung von Textilien und Bekleidung

81.3. Spinnstoffaufbereitung und Spinnerei

81.4. Herstellung von Textilprodukten

81.5. Herstellung von Textilkleidung

82. Wissenschaftsforschung

82.1. Wissenschaftsethik

82.2. Wissenschaftsphilosophie

82.3. Wissenschaftsgeschichte

82.4. Wissenschaftssoziologie

83. Informationserfassung und Datenspeicherung

83.1. Archivierung

83.2. Archivwissenschaft

83.3. Bibliothekservice

83.4. Bibliothekwissenschaft

83.5. Museumstätigkeiten

83.6. Museumskunde

84. Datenverarbeitung und Informationstechnologie

84.1. Informatik (<- Computerwesen)

84.2. Informations- und Kommunikationstechnologie (<- IKT)

84.3. Informationstechnologiedienste

85. Statistik

86. Toxikologie

87. Tabakindustrie

88. Pelzindustrie

88.1. Verarbeitung und Färben von Pelzware

88.2. Herstellung von Pelzwaren

88.3. Herstellung von Pelzkleidung

89. Schutz und Sicherheit

89.1. Detektei-Bereich

89.2. Personenschutz

89.3. Notfallzentralen

89.4. Aufrechterhaltung der Ordnung

89.5. Verkehrssicherheit (<- Straßenverkehrssicherheit)

89.6. Kaminreinigung

89.7. Feuerwehr

89.8. Rettungsdienst

89.9. Polizei

89.10. Grenzsicherung

89.11. Risikomanagement

89.12. Strahlungsschutz

89.13. Geheimdiensttätigkeiten (<- Geheimdienstagenturen)

89.14. Sicherheitskontrollen

89.15. Sicherheitstechnologie

89.16. Arbeitssicherheit (<- Betriebssicherheit; Sicherheit am Arbeitsplatz)

89.17. Justizvollzug

90. Forschung und Entwicklung

90.1. Forschungs- und Entwicklungsinfrastruktur

90.2. Forschungspolitik

90.3. Forschungsunterstützung

91. Arbeitsleben und Arbeitsschutz

91.1. Berufsberatung (<- Berufsberatungsservice)

91.2. Berufsorganisationen

91.3. Gewerkschaften

91.4. Wirtschafts- und Arbeitgeberverbände

91.5. Ergonomie

91.6. Personalmanagement

91.7. Personaltraining

91.8. Bereitstellung von Humanressourcen

91.9. Beschäftigungsaktivitäten

91.10. Arbeitsorganisation

91.11. Arbeitsrecht

91.12. Arbeitshygiene

91.13. Arbeitssicherheit (<- Betriebssicherheit; Sicherheit am Arbeitsplatz)

91.14. Arbeitspolitik

92. Astronomie und Raumfahrtsforschung

92.1. Astrophysik

92.2. Astrometrie

92.3. Kosmologie

92.4. Himmelsmechanik

93. Glaubensvorstellungen und Okkultismus

93.1. Alchemie

93.2. Anthroposophie

93.3. Astrologie

93.4. Wahrsagerei

93.5. Feng-Shui

93.6. Kabbalah

93.7. Magie

93.8. Numerologie

93.9. Parapsychologie

93.10. Spiritualismus

93.11. Theosophie

93.12. Ufologie

94. Religion

94.1. kirchliche und sonstige religiöse Vereinigungen

94.2. Theologie

94.3. Religionsphilosophie

94.4. Religions- und Glaubensrichtungen

94.5. Religionsgeschichte

94.6. Religionsunterricht

94.7. Religionswissenschaft

95. Freizeit und Hobbys

96. Wasserversorgung, Abwasser- und Abfallentsorgung

96.1. Abfallsammlung

96.2. Abfallbehandlung und Beseitigung

96.3. Recycling

96.4. Wasserentnahme, -aufbereitung und –verteilung

96.5. Kanalisierung und Abwasserentsorgung

97. Kommunikation und Informationsübermittlung

97.1. Massenmedien

97.2. Journalismus und Journalistik

97.3. Verlagswesen

97.4. Druck und Reproduktion von Aufzeichnungen

Appendix 3.5 Primary reference sources of TermFactory domain classification

All online resources were accessed in April–August 2010.

1. *Statistical Classification of Economic Activities in the European Community (NACE Rev. 2)*: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl= LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=EN&IntPcKey=&StrLay outCode=HIERARCHIC.
2. *Finnish version of NACE Rev. 2 with comments by Statistics Finland (Tilastokeskus)*: http://www.stat.fi/meta/luokitukset/tieteenala/001-2007/kuvaus.html.
3. *International Standard Industrial Classification of All Economic Activities (ISIC Rev. 4)*: http://unstats.un.org/unsd/cr/registry/isic-4.asp.
4. *General Finnish Thesaurus (YSA)*: http://vesa.lib.helsinki.fi/ysa/index.html
5. *Finnish General Upper Ontology (YSO)*: http://www.yso.fi/onki2.
6. *Eurovoc, the EU's Multilingual Thesaurus*: http://eurovoc.europa.eu.
7. *Universal Decimal Classification (UDC):* http://www.udcc.org.
8. *Abridged Finnish version of UDC*: http://www.kansalliskirjasto.fi/kirjastoala/fennica/ fennica_udkkaavio.html.
9. *Wikipedia, the Free Encyclopedia*: http://www.wikipedia.org.
10. *Helsinki City Library Classification (HKLJ)*: http://hklj.kirjastot.fi/en-GB.
11. *Recommendations Concerning the International Standardization of Statistics on Science and Technology*: http://portal.unesco.org/en/ev.php-URL_ID=13135& URL_DO=DO_TOPIC&URL_SECTION=201.html.
12. *Field of Science and Technology Classification (FOS)*: http://www.stat.fi/meta/ luokitukset/tieteenala/001-2007/kuvaus.html.
13. *Library of Congress Classification (LCC)*: http://id.loc.gov.

Appendix 4   Documentation of sources

1. The following definitions of the basic bibliographic concepts are essential for correct use of the templates in this Appendix:

   - A *monograph* is a bibliographical resource that is complete in one part or intended to be completed in a finite number of parts (ISBD 2011: 270).
   - A *serial* is a continuing resource issued in a succession of discrete issues or parts, usually bearing numbering, which has no predetermined conclusion (ISBD 2011: 276). Examples of serials include printed and electronic journals, newspapers and monographic series.
   - An *integrating resource* is a continuing resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole (ISBD 2011: 267). An updating website or database are examples of integrating resources.
   - A *contribution* is an independent constituent unit of a document (ISO 12615:2004: 2). Articles are a typical example of contributions.

2. The following abbreviations are used in the templates in this Appendix (cf. ISBD 2007: 0-2; ISBD 2011: 10):

   - M = Mandatory. Mandatory elements are required in all situations if applicable.
   - MA = Mandatory if Applicable. Elements with this label are mandatory if the information is available or applicable to the resource.
   - C = Conditional. Conditional elements are required under certain conditions, for example, when necessary for identification or otherwise considered important to the users. If the condition is not met, the use of the element is optional.
   - O = Optional. Optional elements may be included or omitted at the discretion of the contributor.

3. Bolded rows in the templates (e.g., Title, Responsibility, Edition in *Appendix 4.2*) are logical divisions, not data categories.

Appendix 4.1 Basic template for the documentation of written sources

Field	Note	M/C/O
0.1 Language of the bibliographic record	Values should be picked from the language selector.	M
0.2 Script of the bibliographic record	Values should be picked from the script selector.	M
1 Bibliographic citation	Information about the resource provided in the way the contributor wants it to appear in the bibliography. The information shall be sufficient for an unambiguous identification of the resource and shall include the following components: ■ title of the resource (at least the *title proper*) ■ responsibility (at least *primary responsibility*) ■ place and date of publication (for published resources) or creation (for unpublished resources) ■ host document and location within it (for contributions) ■ version or edition (if applicable) ■ type of resource specific data (if applicable).	M
2 Format of resource	A picklist with the following recommended values (cf. ISBD 2007: 0-1): printed text, electronic resource, multimedia resource, moving image, sound recording, still image, cartographic resource, notated music resource, other resource (please specify).	M
3 Accessibility	Information on whether the resource is published or unpublished, how it can be accessed and when it was retrieved.	M
4 Supplementary information	Additional relevant information about the resource that does not belong in the previous three data categories.	C
5 Bibliographic standard	Reference to the bibliographic standard used in documenting the resource. A combination of a picklist with predefined values and a free-form extension field may be recommended. The picklist may be updated by moderators based on the feedback from the community.	O

6 Compliance with the bibliographic standard	Indication whether the bibliographic record is fully or partially compliant with the specified bibliographic standard.	MA
7 Source ID	An automatically generated unique and persistent identifier used for unambiguous linking between the bibliographical record and references to it.	M
8 Visual source identifier	A unique, relatively short and transparent alias intended to be used as a reference to the bibliographic record in the entries.	M

Appendix 4.2 Extended template for the documentation of written sources

Field	Note	M/C/O
0 Administrative data		
0.1 Language of the bibliographic record	Values should be picked from the language selector.	M
0.2 Script of the bibliographic record	Values should be picked from the script selector.	M
1 Title	A word or phrase, or a group of characters, usually appearing on a resource, which is the name of the resource (ISBD 2007: E-13).	
1.1 Title proper	The main title of a resource in the form in which it appears on the prescribed sources of information for the title and statement of responsibility area (ISBD 2007: E-14).	M
1.2 Parallel title	A title presented on the prescribed sources of information for the resource as an equivalent in another language and/or script of the title proper (ISBD 2007: E-9).	C
1.3 Other title information	Information appearing in conjunction with and subordinate to the title proper or parallel title(s) (ISBD 2007: 1.4-1).	C
2 Responsibility	Statement of responsibility consists of name(s), phrase(s), or group(s) of characters relating to the identification and/or function of any persons or corporate bodies responsible for or contributing to the creation or realization of the intellectual or artistic content of a work contained in the resource described (ISBD 2007: 1.5-1).	
2.1 Primary responsibility	Name of person(s) or organization(s) carrying the principal responsibility for the resource, such as the author(s) and/or responsible organization(s).	M
2.2 Secondary responsibility	Name of person(s) or organization(s) carrying secondary responsibility for the resource, such as translator(s).	C
3 Edition	Edition means all copies of a resource produced from substantially the same original input and issued by the same agency or person (ISBD 2007: E-3).	

232

3.1 Edition	E.g., 2nd edition, new edition, revised edition, standard edition, large print edition, fourth state, May 1970 script (ISBD 2007: 2.1-1).	MA
3.2 Parallel edition	Equivalent of the edition statement in another language and/or script (ISBD 2007: 2.2-1).	O
3.3 Primary responsibility relating to edition	Name(s) of person(s) or organization(s) carrying the principal responsibility for the edition, such as the author(s) and/or responsible organization(s).	MA
3.4 Secondary responsibility relating to edition	Name(s) of person(s) or organization(s) carrying secondary responsibility for the edition, such as translator(s).	C
3.5 Other edition information	For example, information about the version and date of the last update for online resources.	C
4 Type of resource specific data		
4.1 Type of resource specific description	For example, for patent documents, type of resource specific description may include application identifier, date of submission of the application and country of issuing office.	MA
5 Publication / production		
5.1 Place of publication, production and/or distribution	Name of the place associated in the prescribed source of information with the name of the principal publisher, producer and/or distributor. If no publisher, producer or distributor is named, it is the place from which the resource was issued or distributed (ISBD 2007: 4.1-1). When more than one place is associated with the name of a single publisher, producer or distributor, the place made more prominent by typography or the place name that appears first is given (ISBD 2007: 4.1-2). Second and subsequent places may be omitted and replaced with [etc.] or its equivalent in another language or script: Wien [etc.] (ISBD 2007: 4.1-3). For online resources, the place of publication is usually the URI.	M
5.2 Name of publisher, producer and/or distributor	Person(s) or organization(s) that effect the publication, production and/or distribution or release activities for the resource (ISBD 2007: 4.2-1).	M

5.3 Date of publication, production and/or distribution	This is usually the year of publication, production or distribution. For the full dates, the recommended format is the ISO 8601 extended format (YYYY-MM-DD).	M
6 Physical characteristics		
6.1 Type of medium	Material in any format that carries and communicates information content (ISBD 2007: E-7), e.g., book, CD-ROM, or DVD.	C
6.2 Format	The manner in which data is arranged in a medium of input, output or storage (ISBD 2007: E-5). Suggested values (cf. ISBD 2007: 0-1): printed text, electronic resource, multimedia resource, moving image, sound recording, still image, cartographic resource, notated music resource, other resource (please specify).	M
6.3 Extent	E.g., the number of pages.	C
7 Host		
7.1 Reference to host (for multipart resources and contributions)	Reference to the source ID of the host publication.	MA
7.2 Part number (for multipart resources)		C
7.3 Location within host (for contributions)	E.g., issue, volume, chapter number, pagination.	M
8 Series		
8.1 Reference to series (for parts of monographic series and individual issues of serials)	Reference to the source ID of the series.	MA
8.2 Volume number (for parts of monographic series and individual volumes of serials)		MA

8.3 Issue number (for individual issues of serials)		MA
8.4 Numbering (for serial as a whole)	Numbers and/or dates of coverage of the first and/or last issue of the serial (ISBD 2007: 3.3-1).	C
9 Identifier		
9.1 ISBN (for monographs)	A unique identifier assigned by the national ISBN agency and based on ISO 2108 (ISBD 2007: E-6).	C
9.3 ISSN (for serials)	An identifier assigned by the ISSN Network and based on ISO 3297 (ISBD 2007: E-6).	C
10 Accessibility		
10.1 Accessibility of the source	Information on whether the source is published or unpublished.	M
10.2 Access requirements	Technical requirements (e.g., hardware and system requirements for electronic resources) and/or the terms upon which the resource is available (e.g., subscription to a service).	C
10.3 Date of citation (for online and other integrating resources)	The recommended format is the ISO 8601 extended format (YYYY-MM-DD).	MA
10 Notes		
10.1 Note		O

Appendix 4.3 Template for the documentation of private sources for bibliographic purposes

Field	Note	M/C/O
0.1 Language of the record	Values should be picked from the language selector.	M
0.2 Script of the record	Values should be picked from the script selector.	M
1 Formatted name	Full name provided in the way it should appear in the bibliography. The name may also be accompanied by desired honorific prefixes, suffixes and titles and contain elements of inline formatting (e.g., italics).	M
2 Family name		M
3 Given name		M
4 Job title		M
5 Role in organization	Role, occupation, or business category within the organization.	C
6 Name of organization		M
7 Organizational unit		C
8 Other titles	E.g., academic titles.	C
9 Note		O

Appendix 4.4 Template for the documentation of private sources for contact purposes

Field	Note	M/C/O
0 Administrative data		
0.1 Language of the record	Values should be picked from the language selector.	M
1 Personal data		
1.1 Formatted name	Full name with desired honorific prefixes, suffixes and titles. The field may also contain elements of inline formatting (e.g., italics).	M
1.2 Family name		M
1.3 Given name		M
1.4 Additional name		C
1.5 Name prefix		O
1.6 Name suffix		O
1.7 Birthday	Recommended format is the ISO 8601 extended format (YYYY-MM-DD).	O
2 Photograph		
2.1 Photograph encoding	Encoding of the binary data, e.g., BASE64.	C
2.2 Photograph format type	Graphics format for the photograph, e.g., JPEG.	C
2.3 Link to image	Link to the graphical file.	C
3 Titles and affiliations		
3.1 Job title		M
3.2 Role in organization	Role, occupation, or business category within the organization.	C
3.3 Name of organization		M
3.4 Organizational unit	Name of the unit(s) of the organization associated with the person.	C

3.5 Agent	Link to another record that provides information about the person who acts on behalf of the individual in question (e.g., administrator, assistant or secretary).	O
3.6 Other titles	E.g., academic titles.	C
3.7 Note related to titles and affiliations		O
4 Language(s)	Values should be picked from the language selector.	
4.1 Native language		C
4.2 Language of written communication		M
4.3 Language of oral communication		C
4.4 Note related to language of communication		O
5 Time and location		
5.1 Domestic country of the consultant	Recommended format is ISO 3166-1 (see http://www.iso.org/iso/country_codes/iso_3166_code_lists/english_country_names_and_code_elements.htm, accessed 1.3.2013).	C
5.2 Location of the consultant	The best practice is to use generally acknowledged standards and ontologies of geographical places and to provide the name in the official languages of the country in question.	C
5.3 Time zone of the consultant	Offset from Coordinated Universal Time (UTC) specified as a positive or negative difference in units of hours and minutes (+hhmm or -hhmm). See ISO 8601:2004 for more information.	C
5.4 Consultation date(s)	Time span(s) when private consultations were given. Recommended format is YYYY.MM.DD-YYYY.MM.DD or shorter versions, e.g., YYYY.	M
5.5 Note related to time and location		O
6 Delivery address		
6.1 Full delivery address of the consultant		C

6.2 Delivery address type	E.g., domestic, international, postal, parcel, home, work.	C
6.3 Note related to delivery address		O
7 Telephone		
7.1 Full telephone number for domestic calls	Telephone number with all the prefixes required for making a domestic call.	C
7.2 Full telephone number for international calls	Telephone number with all the prefixes required for making an international call.	
7.3 Telephone type	Type of telephone associated with the telephone number of the person (such as Home, Work, Mobile).	C
7.4 Preferred telephone number	This field specifies which telephone number is the preferred one if several are provided.	O
7.5 Note related to communication over the telephone		O
8 E-mail		
8.1 E-mail address		C
8.2 Preferred e-mail address		C
8.3 Note related to communication over e-mail		O
9 Instant messaging		
9.1 Instant messaging agent	E.g., ICQ.	C
9.2 Preferred instant messaging agent		C
9.3 User ID in instant messaging system		C
9.4 Note related to communication by instant messaging		O

10 Voice/video calls		
10.1 Voice/video call agent	E.g., Skype.	C
10.2 Preferred voice/video call agent		C
10.3 User ID in voice/video call agent		C
10.4 Note related to voice/video calling		O
11 Webpage		
11.1 Parent site		C
11.2 Homepage		C
11.3 Note related to webpage		O
12 Social media applications		
12.1 Social media agent	E.g., Facebook, LinkedIn, Twitter.	C
12.2 Preferred social media agent		C
12.3 User ID in social media agent		C
12.4 Note related to communication via social media.		O

Appendix 5   Documentation of contributors' expertise

Field	Expected value
0 Administrative data	
0.1 Language of the record	value picked from the language selector
1 General language expertise	
1.1 Language	value picked from the language selector
1.2 Language proficiency in written communication in the specified language	level of proficiency according to the *Common European Framework of Reference for Languages* (CEFR): A1 Breakthrough or beginner A2 Waystage or elementary B1 Threshold or intermediate B2 Vantage or upper intermediate C1 Effective operational proficiency or advanced C2 Mastery or proficiency
1.3 Theoretical language expertise	
1.3.1 dissertation	title, university and year of approval
1.3.2 post-graduate studies	university and time period
1.3.3 final-year thesis	title, university and year of approval
1.3.4 undergraduate work	title, university and time period
1.3.5 notes	
1.4 Language education	
1.4.1 university or other high school	organization and number of credits
1.4.2 professional education institution	organization and number of credits
1.4.3 other courses	organization and number of credits
1.4.4 notes	
1.5 Practical experience	
1.5.1 translation	direction, volume, time period and customer/employer

1.5.2 technical writing	volume, time period and customer/employer
1.5.3 editing	volume, time period and customer/employer
1.5.4 language revision	volume, time period and customer/employer
1.5.5 language teaching	extent, time period and customer/employer
1.5.6 other (please, specify)	
1.5.7 notes	
2 Domain expertise	
2.1 Domain	value picked from the domain selector and possibly supplemented by a free-form specification
2.2 Theoretical expertise	
2.2.1 dissertation	title, university and year of approval
2.2.2 post-graduate studies	university and time period
2.2.3 final-year thesis	title, university and year of approval
2.2.4 undergraduate work	title, university and time period
2.2.5 notes	
2.3 Education	
2.3.1 university or other high school	organization and number of credits
2.3.2 professional education institution	organization and number of credits
2.3.3 other courses	organization and number of credits
2.3.4 notes	
2.4. Practical experience	volume, time period and customer/employer
3 LSP expertise	this section may be filled in mulitple times for various combinations of languages
3.1 Domain	value picked from the domain selector and possibly supplemented by a free-form specification

3.2 Language 1	value picked from the language selector
3.3 Language 2	value picked from the language selector
3.4 Text production expertise in the specified domain in Language 1	from 1 to 5; 1 = basic, 5 = excellent
3.5 Translation expertise from Language 1 to Language 2	from 1 to 5; 1 = basic, 5 = excellent
3.6 Translation expertise from Language 2 to Language 1	from 1 to 5; 1 = basic, 5 = excellent
4 Expertise in terminology and terminography	
4.1 Theoretical expertise in terminology and terminography	
4.1.1 dissertation	title, university and year of approval
4.1.2 post-graduate studies	university and time period
4.1.3 final-year thesis	title, university and year of approval
4.1.4 undergraduate work	title, university and time period
4.1.5 notes	
4.2 Education in the field of terminology and terminography	
4.2.1 university or other high school	organization and number of credits
4.2.2 professional education institution	organization and number of credits
4.2.3 other courses	organization and number of credits
4.2.4 notes	
4.3 Practical experience in terminology and terminography	
4.3.1 participation in compilation of a published dictionary or glossary	bibliographic description, name and time span of the project, role in the project, responsible organization
4.3.2 participation in compilation of an unpublished dictionary or glossary	description of the dictionary or glossary, name and time span of the project, role in the project, responsible organization
4.3.3 occasional assignments involving terminology work	nature of the assignments, time period, customer/employer
4.3.4 notes	

5 Experience in online collaborative content creation	
5.1 project	resource, time period, role
5.2 notes	

Publications from the *TermFactory* project

1. Kudashev, I. (2009). Improving Compatibility of Terminological Collections with a Bridging Classification of Data Categories. *Terminologija*. Vol. 16. 36–55.
2. Kudashev, I. (2010). What can be an Object of Terminological Description in a Term Bank? In Nissilä, N. and Siponkoski, N. (eds) *Language and Emotions. Theory of Translation, LSP and Multilingualism. VAKKI symposium XXX. Vaasa, 12–13.2.2010*. Vaasa: University of Vaasa. 142–152.
3. Kudashev, I. and Kudasheva, I. (2010). Semiotic Triangle Revisited for the Purposes of Ontology-Based Terminology Management. In *TOTh 2010. Actes de la quatrième conférence TOTh – Annecy – 3 & 4 juin 2010.* Annecy: Institut Porphyre. 83–100.
4. Kudashev, I., Kudasheva, I. and Carlson, L. (2010). TermFactory: A Platform for Collaborative Ontology-Based Terminology Work. In Dykstra, A. and Schoonheim, T. (eds) *Proceedings of the XIV Euralex International Congress (Leeuwarden, 6–10 July 2010)*. Ljouwert: Fryske Akademy; Afûk. 931–936.
5. Kudashev, I., Carlson L. and Kudasheva, I. (2010).TermFactory: Collaborative Editing of Term Ontologies. In Bhreathnach, Ú. and Barra-Cusack, F. (eds). *Terminology and Knowledge Engineering Conference 2010. Presenting Terminology and Knowledge Engineering Recourses Online: Models and Challenges*. Dublin: Fiontar, Dublin City University. 479–500.
6. Norta, A., Yangarber, R. and Carlson, L. (2010). Utility Evaluation of Tools for Collaborative Development and Maintenance of Ontologies. In *EDOCW '10: Proceedings of the 2010 14th IEEE International Enterprise Distributed Object Computing Conference Workshops, Vitoria, Brazil, 25–29.10.2010.* Washington: IEEE Computer Society. 207–214.
7. Kudashev, I. (2011). Documentation of Sources in Terminology Management Systems, with Particular Emphasis on Collaborative Platforms. In Madsen, B.N. and Thomsen, H.E. (eds). *Ontologier og taksonomier. Rapport fra NORDTERM 2009, København, Danmark, 9.–12. juni 2009*. Copenhagen: Copenhagen School of Business. 226–235.
8. Kudashev, I. (2011). Ob ob'ekte terminologičeskogo opisaniâ v terminologičeskom banke dannyh. In *Terminologia i znanie. Materialy II Meždunarodnogo simpoziuma (Moskva, 21–22 maâ 2010 g.)* / Ed. by S.D. Šelov. Moskva: Azbukovnik. 203–214. (= On the Object of Terminological Description in a Term Bank).

9. Kudashev, I. (2011). Domain Classification for Terminology Management: Principles of Compilation. In Nissilä, N. and Siponkoski, N. (eds). *LSP, Translation and Multilingualism: VAKKI symposium XXXI, Vaasa, 11–12.2.2011*. Vaasa: University of Vaasa. 176–187.

10. Kudashev, Igor. (2012). Workflow Models for Collaborative Terminology Work. In Ylisalmi, A. (ed). *Samarbetet ger resultat: från begreppskaos till överenskomna termer. Rapport från Nordterm 2011, Vasa, Finland, 7–10 juni 2011*. Helsingfors: Nordterm; TSK. 130–136.

11. Kudashev, I. and Carlson, L. (2012). Administrative Data in Terminological Collections and Term Ontologies. In Aguado de Cea et al. (eds). *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012), 19–22 June 2012, Madrid, Spain*. Madrid: TKE organizers. 270–282.

12. Kudashev, I. (2012). International Medical Classifications from the Terminological Point of View (cases: *International Classification of Diseases* and *Terminologia Anatomica*). In Nissilä, N. and Siponkoski, N. (eds). *Languages in Motion. VAKKI symposium XXXII. Vaasa 10–11.2.2012*. Vaasa: University of Vaasa. 118–129.

13. Kudashev I.S. (2013). Quality Assurance Infrastructure for Term Banks and Collaborative Terminological Projects. In *Terminologia i znanie: Materialy III Meždunarodnogo simpoziuma* / Ed. by S.D. Šelov. Moskva: IRÂ RAN. 160–168.