Terminology for Large Organizations

Terminology Starter Guide

Contents

About

Developed by:

Established in 2011, Terminology for Large Organizations (TerminOrgs) is a consortium of terminologists and other communications professionals who promote terminology management as an essential communications strategy in large organizations.

This group is a forum to discuss and develop guidelines and best practices for large-scale terminology management. Our mission is to raise awareness about the role of terminology for effective communications, knowledge transfer, education, risk mitigation, content management, translation and global marketing, with a focus on large organizations. TerminOrgs represents stakeholders of terminology standards and tools. We work to determine and promote the economic value of managing terminology.

To learn more about TerminOrgs and our activities, please visit "www.terminorgs.net"

Contributors

Petr Oklestek (Avaya), Hanne Smaadahl (SAP), Christine Hug (Government of Canada Translation Bureau), Diahnn Theophilus (Xerox), Dino Ferrari (CLS Communication), Kara Warburton (Termologic), Michael Wetzel (ESTeam), Sue Kocher (SAS), Vicki Leary (SAS)

Version

December 2012. This document is updated periodically. Please visit "http://www.terminorgs.net" to download the latest version.

Copyright

Who should read this document?

The target audiences are teams or individual members of departments in big corporations, government or language service providers that need to manage terminology. It is intended as an introduction to teams and individuals who are new to terminology, and who are not sure whether or how to get started with terminology management. This guide provides readers with an overview of the topics and points them to further resources. It is not intended as an exhaustive manual.

Introduction

3.1 What is a term?

A term can be a single word (simple term), a multi-word expression (complex term), a symbol (©, ®) or a formula (H_2O) that designates a particular concept within a given subject field ([1]). Terms may coincide with words in general language; however, it is their context, or usage in a specialized subject field, that determines their precise meaning. A term may contain symbols and can have variants, such as abbreviations or different forms of spelling.

Although terms can belong to any of the open word classes (nouns, verbs, adjectives, adverbs), in reality 80% or more in any given field are nouns. Furthermore, in a terminology management system (TMS) aimed at translation, adjectives and adverbs are often not included in the TMS as their meaning can change slightly depending on the context in which they are used, and therefore, it can be difficult to determine target language equivalents for a fixed terminology entry.

Nowadays, terms are typically recorded and managed in a terminology database (termbase).

Term entries in a termbase frequently describe words with a specialized meaning (technical, scientific, market-specific, political, etc.) and words that are prominent in customer-facing materials (slogans, names of products, features, programs, parts, frequently-occurring words, labels on user interfaces, packaging, etc.).

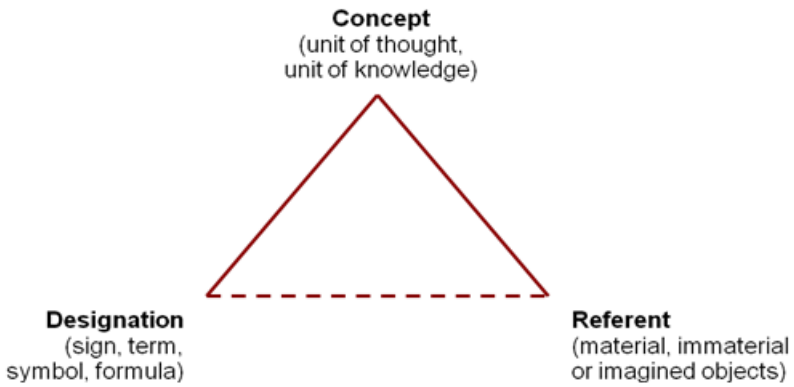
The traditional view places terms that convey special meanings as opposed to “words” that convey general meanings. However in commercial environments, any word or expression that is deemed to be important for the company’s business and communications qualifies as a “term” subject to inclusion in the company termbase, regardless of whether its meaning is “special” or not. Thus, in commercial environments, even words that do not have a specialized meaning may be considered to be terms as they also need to be proactively managed: terms that occur frequently in company materials need to be used and translated consistently.

---

1. The Pavel Terminology Tutorial, Translation Bureau of Canada

3.2 What is a concept?

A concept is a unit of knowledge created by a unique combination of characteristics (ISO 1087:1-2000). A concept is a discrete unit of thought that can refer to something material, immaterial or even imagined. Concepts are the units within terminology management to which terms are designated.



3.3 What is terminology?

“Terminology” has two meanings. First, terminology refers to a set of terms in a specialized area, such as “networking terminology” or “automobile manufacturing terminology”. Second, it is the name of an academic and professional discipline associated with studying and managing terms. Considered a branch of linguistics, terminology is closely related to lexicology (defining words and creating dictionaries), but with a focus on concepts (analysis, definition, denotation) in special domains. The field of terminology typically supports content creation, translation and other forms of knowledge management.

3.4 What is terminology management?

Terminology management is the set of activities carried out to ensure that the correct terms are used consistently in all company materials, in support of end-to-end product development, communication, translation/localization, and distribution. First, it involves collecting the terms in a domain, and identifying and eliminating inconsistencies, including identifying synonyms and abbreviations, and controlling their use. Second, it encompasses documenting these terms with the appropriate metadata, such as definitions, subject field, and part of speech. Terminology management also comprises the distribution and dissemination of terms as reference materials for writers and translators and as linguistic resources for other systems including, but not limited to, content management and authoring tools, translation tools, product taxonomies, and search optimization.

3.5 What is a terminology management system?

A terminology management system (TMS) is computer software that helps you to store and retrieve terminological information.

It stores terms, associated and illustrative information such as definition, context, or an image, classificatory information such as domain or usage recommendation, as well as administrative information such as creator, creation date, and change history. Some terminology management systems have a fixed set of fields; others are customizable for the needs of users and user groups.

A TMS retrieves terms by means of search and filter features. A search might look for a string within terms only, or in other text fields in the database; the latter is known as a full-text search. Searches are carried out directly in the TMS, i.e. within its own user interface (browser or rich client application). Some systems are also integrated into other language technology applications, such as translation editors or authoring tools. These tools may then also execute searches in the TMS and display the results within their own interface.

3.6 Role of terminology management in corporate activities

Terminology management is the systematic research, documentation, storage and distribution of concepts along with the terms that denote those concepts. In a corporate environment, terminology management needs to be a proactive, multi-disciplinary and integrated process in order to be successful. Proactive terminology management aspires to intercept terminology at the time of inception, when concepts are created and named. It also means that various people in the organization participate in terminology discussions, decisions and implementations from R&D to marketing, from developers to product managers, from technical writers and editors to translators, and from sales force to support

teams. Finally it means that terminology management is an integral part of the business' workflows and is built into formal objectives and performance measurements of the key stakeholders and users, such as technical writers and translators.

Rather than limiting terminology management to an isolated activity in the translation department, the successful organization incorporates it as a holistic approach that unifies corporate activities such as design, quality control, branding and marketing, search, text mining and analysis, content management and controlled authoring, as well as human and machine translation.

3.7 Role of terminology management in government

Good terminology management in government has some of the same objectives as in product localization or commercial communications: reducing translation time and assisting translators in producing consistent, high quality translations. However, even before the translation step, good terminology management also helps public servants write better, clearer documents in the first place.

Wasted time is wasted taxpayer money and a government should be clear and coherent when communicating internally and with its citizens. For governments, particularly those with more than one official language, consistent terminology is vital - when writing laws, for example. Imagine what could happen if the terminology used in a country's constitution was inconsistent within one language and then translated inconsistently into others, with each version having the same legal weight. One arm of a government cannot accurately and effectively communicate with another without consistent terminology. Governments cannot effectively deliver services to their "clients" (the citizenry) without consistent terminology. Multiply the number of misunderstandings (including potentially grave ones) that can be made about one concept in one language by the number of official languages and it is clear that terminology management is important in government. Over time, poor translations could give the impression of carelessness, and even result in alienation of voters who speak a particular official language. Managing terminology well is in the best interests of governments everywhere, easing internal communication and also helping to present a unified message to other countries and international organizations. Some governments that have realized this (Switzerland, France and Canada, among others) have dedicated resources to terminology management.

In countries that have more than one official language, it is a citizen's right to receive information in the official language of his or her choice. Terminology management is therefore a key component in a government's legal responsibility to provide linguistic services. Terminology management also serves a role in support and preservation of endangered languages

3.8 Role of terminology management in technical writing

In many corporate environments that support terminology management, it is the technical writer who is on the forefront of terminology development. In this environment, the technical writer, the product manager, and the terminologist identify new terminology that emerges as products are being developed: product user interfaces, error and system messages, internal product design specifications, vocabulary

to be shared in everyday communication, and customer documentation. The terminologist adds the new terms to the company termbase. Of ultimate importance is agreement on terms and definitions for consistent use in all corporate collateral. The obvious beneficiary of such due diligence is the customer.

The technical writer must be aware of emergent terminology during the early phase of a product's development cycle. A clear identification and understanding of new terms is critical to the writer who develops detailed conceptual topics in the documentation.

The writer is often an active member of a product development team, working with R&D to develop clear and consistent terminology within and across product teams in the corporation. The writer must be sufficiently skilled to recognize the difference between everyday language words (terms to be omitted from consideration) and legitimate, "new" terminology to be researched and documented.

In some environments, the writer participates in the process of entering and managing product-specific terminological data in the corporate termbase using a particular TMS. The writer's level of access to particular dictionaries and sections in the corporate termbase might vary according to the organization's access policies. For example, the writer might have write access to a section of the termbase to enter a set of terms and definitions with a status of "unprocessed." Typically, either the terminologist or the editor has the authority to assign a "finalized" (approved) status to terms and definitions, at the conclusion of the specified review process.

In many companies the writer works directly with localization specialists, supplying text and contextual information about terminology for translation.

The final stage in the writer's work flow could include the generation of a glossary to be incorporated in the final document deliverable. Using a glossary extraction function in the TMS, the writer requests a production-quality glossary of finalized terms and definitions for the specified project. The final glossary is included in the document for delivery to the customer. Furthermore, members of the product R&D community, as well as all other employees and relevant parties, should be able to browse and filter the termbase for terminology.

3.9 Role of terminology management in localisation

Good terminology management reduces translation time and assists translators in producing good quality and consistent translations. All this helps to significantly reduce localization costs; translators can spend less time researching or querying terminology and quality issues are minimised. Consistency in target terminology enables previously localized material to be reused; text from different but similar material can be successfully merged to produce a new document. A concept can be represented by many different terms in most languages. Bear in mind that during a localization project, concepts usually need to be represented in multiple languages – so standardizing terminology in both source and target terminology facilitates accuracy and establishes a consistent approach. Reducing the need for rework by getting it right first time, i.e. by building a termbase, is not only an exercise in accuracy or an aid to establishing a good brand image, but it also saves money in the long-term.

What are the benefits of terminology and terminology management?

Managing corporate terminology provides benefits to the bottom line, to customer relations and the overall image of an organization.

4.1 Reducing costs and time to market

Well managed terminology increases the ability to reuse and retrieve information. Editors, translators and other specialists spend less time searching for correct terms, definitions and equivalents, hence increasing their efficiency. Furthermore, the consistent use of terminology throughout the document life cycle leads to a reduction of errors and their propagation in translated versions, which in turn reduces the cost of revising and reprinting documents and helps shorten time to market. After-sales costs can be reduced if misunderstandings or errors are avoided by producing comprehensible documents, including instruction manuals.

4.2 Improving quality

Managed corporate terminology improves the quality of products and corporate documents. This in turn increases the usability of products and thereby customer satisfaction which could lead to increased market share. Technical texts are easier to understand if the terms are used consistently in all official languages. Documents that are intended for clients will profit from a selection of terms according to target group, and from the definition of specialised expressions, as well as avoidance of confusing jargon, abbreviations and terms.

4.3 Strengthening brands and protecting intellectual property

Successful brands create a positive impression through positive customer experience, which in turn leads to customer loyalty. High-quality documents are a key contributor, and terminology contributes considerably to the success of corporate brands through clear and successful communication. In the

era of global one-voice strategies and virtualisation of client contacts, professional management of terminology in all markets and languages is all the more important.

4.4 Preserving know-how

Concepts are considered to be the basic units of knowledge. Terms are the linguistic representation of concepts, whereas definitions can be seen as standardized explanations of concepts. These two elements, terms and definitions, are key components in termbases and therefore tremendously precious to a company. With every specialist who leaves the company, a portion of knowledge is lost. This know-how can be retained if it is stored in a termbase. Human Resources and senior employees can use termbases (when available) to explain basic concepts to new employees, thus making sure that continuity of knowledge is preserved.

Dealing with terms

5.1 What terms should be defined?

Preparing definitions is time-consuming, and it might not be practical to write a definition for each term in your termbase. Therefore, you need to decide which terms need definitions, and which terms are self-explanatory. You should define terms that are not found in general dictionaries, technical terms (terms that have a special meaning in a field), the acronyms, abbreviations and initialisms that your people need to know, official proper nouns (like product names or the official name of a government entity), terms that have caused confusion for you in the past, and terms for which there is a preferred or a forbidden synonym.

If the primary stakeholders of your termbase are translators, the needs of translators with respect to definitions should be a key determining factor. For instance, if a term has multiple meanings in the company (homograph), then each meaning should be clearly defined since each distinct usage of the term is in fact a unique concept and is likely to have its own translation. If a term is similar in meaning to another term, the two should be clearly defined so that the nuance between the two can be properly reflected in the translations. Whenever there is potential confusion by translators about similar terms, or terms with many meanings, these are terms that should be defined.

If another stakeholder of your termbase is customers, that is, when the termbase is used to generate glossaries as product information, then you need to define terms with which customers may be unfamiliar. Focus on the frequently-occurring product-specific terms such as the names of product features and underlying technologies. Ensure that all acronyms are expanded to their full forms and defined.

The choice of terms to define also depends on your goal. For example, the terms you would choose to define if your goal is to make your company communicate better internally will differ from those you would choose to define for research in a particular field or for a particular educational purpose.

Ideally, you should have a definition for each term so that anyone consulting a record can see what a concept means. When possible, and if the termbase data structure enables it, add a field to show the context in which a term is used> It can also be helpful to have a field in which explanatory, grammar, and usage information can be provided. Terminology databases without definitions or contexts for the concepts in them are not much better than simple lists of terms, the value and reusability of which diminishes quickly over time and across products.

5.2 Definition (characteristics of a good definition)

A definition should answer the question “what is it?” A good definition specifies the essential and delimiting features of the concept, so that you can tell what it is, as well as how it differs from other related concepts so that the definition unambiguously delineates one concept from another.

When writing a definition, start it with an anchor word that refers to the class the concept belongs to, usually a generic or superordinate word of the same part of speech as the term. (Ex: skirt: An article of clothing that…). Then, add delimiting characteristics (Ex: An article of clothing that hangs from the waist or hips and covers the body below the torso.)

A good definition states what the concept is, not what it is not, nor does it present superfluous information (for example, the name of the inventor of an object is not essential to understanding what the invented object is) or usage details that resemble a Help topic . Describe the term as opposed to paraphrasing it, (for example , meat cleaver: A cleaver for meat – still does not clarify what it is!). Good definitions are non-circular: do not include the term in the definition. Ideally, a definition is simple, concise (one sentence if possible) and contains only essential information, presented clearly. The definition should explain the concept in such a way that it is differentiated from others. The quality of most terminological products will be determined by the quality of the definitions.[1]

---

1. ISO 704:2000: Terminology work - Principles and methods.

5.3 Length of a definition (too long/too short)

Too long or too short is not a matter of word counting, but rather of what people need to know to understand a concept. A definition is too short if it does not contain enough information for someone to understand what concept the term refers to. If you defined a widower as an unmarried male, you would be omitting an essential characteristic that makes him a widower: the fact that he once was married, but that his spouse died. Without that element, the definition would be incomplete. (Not all information about a term should go into the definition; notes on the use of the term or examples of the term in use can be included elsewhere in the entry if required.)

However, a definition is too long if you can take out part of the definition and still grasp what the term represents. Consider your termbase users: they often do not have time to read a definition that spans paragraphs, or to figure out an ambiguous or confusing definition. Write concise definitions that include the essential characteristics of a concept in a definition, and record other useful information about the term in other fields so it is not lost.

5.4 What is the difference between a definition and a description?

A definition limits itself to essential and delimiting characteristics, answering the question, “What is it?” A description can include other more encyclopaedic information like the time and manner of use of the thing described, or the name of its inventor. A description could answer questions such as “Who invented it? How is it used, and by whom?”

Some people wonder if it’s acceptable to add a description or context instead of a definition. (Context is a sample sentence containing the term.) Definitions are the most important textual support, but in the absence of a definition, a description or carefully selected example of the use of the term in context can also provide valuable information to the user. You can have both a context and a definition for a term. A definition alone is the minimum best practice, but context alone is still better than nothing. Ensure that definitions, descriptions, and contexts are recorded in their own dedicated fields in the termbase, that is, do not put descriptions or contexts in the definition field, for example.

Termbase and termbase management

6.1 What is a termbase?

A termbase is a database comprising information about special language concepts and terms designated to represent these concepts, along with associated conceptual, term-related, and administrative information (ISO/DIS 30042:2008, 3.22).

Termbase management is a combination of terminology work and database administrative tasks supporting the systematic collection, description, processing, presentation, and distribution of concepts and their designations.

6.2 Concept-based approach

A concept-based approach to terminology management requires all terminological information pertaining to one concept to be handled as a single terminological entry. In a concept-based system, data is organized around the meaning (the idea or object, the cognitive unit) rather than the form (the term). All terms, variants, and translations that designate a single concept, as well as all descriptive and administrative data belonging to that same concept, are stored in a single terminological entry.

6.3 Data categories

Data categories guide the organization and arrangement of information in each entry of a termbase. The data categories define and govern what type of information goes into each terminological entry, what information is mandatory vs. optional, the format of the information (i.e. free text, predefined pick lists, or Boolean values), and the input method (e.g. manual vs. system generated). A data category is an elementary descriptor in a linguistic structure or an annotation scheme, such as “subject field”, “language”, and “part of speech” (ISO 1087-2:2000). Therefore, data categories often correspond to individual fields in a termbase, such as fields for the part of speech, for the definition, and for other metadata.

6.4 TBX-Basic exchange standard

Term Base eXchange (TBX) is the open, XML-based standard for exchanging structured terminological data. TBX has been approved as an international standard by LISA and ISO (ISO 30042) and is available from www.ttt.org/oscarstandards/.

A TBX variant called TBX-Basic is also available from www.ttt.org/oscarstandards/. TBX-Basic is a lighter version of TBX, and is intended for small or medium-sized organizations. It is well suited for any language application that requires a simple approach to terminology management.

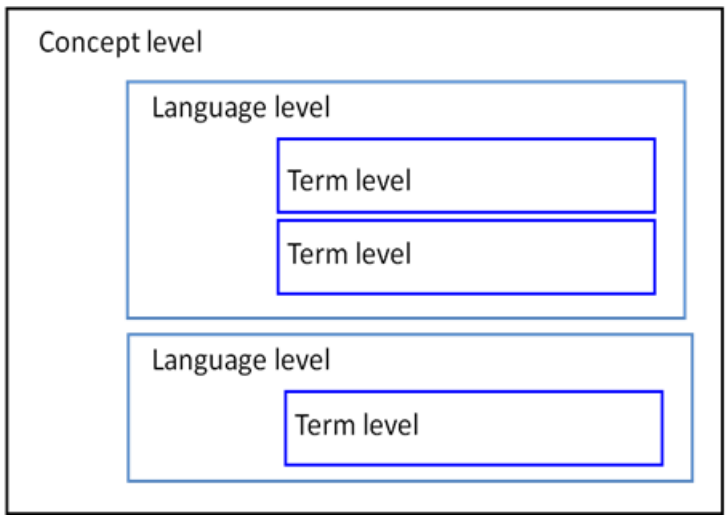
6.5 Common terminology metadata

An inventory of data categories for recording terminology has been defined by ISO Technical Committee 37. The Data Category Registry, ISOcat (http://www.isocat.org/), is a free online service for specification and management of data categories and data category selections for language resources. Anyone can access it and retrieve public data categories and data category selections. In addition, users who register can create and/or share their own data categories and data category selections.

The data categories in TBX-Basic are the most popular ones used in the localization industry, as determined by studies conducted by the Localization Industry Standards Association in 2001 and 2005. (Since the Localization Industry Standards Association (LISA) was disbanded in March 2011 these reports are no longer available; however, the findings from these surveys were incorporated in the first version of TBX-Basic.)

6.6 Terminology entry structure

The TBX-Basic specification outlines the most important terminology data categories and a basic entry structure compliant with TBX and ISO 16642 (Terminological Markup Framework). The entry structure comprises three hierarchical sections: concept level, language level, and term level.


Concept level
Language level
Term level
Term level
Language level
Term level

Key characteristics of a terminology management system

Successful terminology management requires the use of a terminology management system (TMS) that can adapt to constant product development, tools and infrastructure improvements, and organizational changes. The key principles that support these requirements and that should guide the selection of a TMS are summarized in the following sections.

7.1 Data management

A well designed termbase adheres to the following data management principles:

- Single repository
- Concept orientation
- Term autonomy
- Data elementarity

7.1.1 Single repository

All data is managed in a single database or repository. Different types of terminological units (such as single and multi-word terms, acronyms and symbols) are identified using data categories (metadata) and users can access the sections of the terminological database that are relevant to them, based on criteria mapped to the data categories. If a single repository is not possible, then the TMS should allow for some kind of simultaneous search and access to all databases.

7.1.2 Concept orientation

Terminological organization is centred on the meaning – the idea or object as a cognitive unit. In practical terms, this means that homographs (words that are spelt the same but have two or more meanings)

are treated as separate entries, while synonyms are all kept together with the concept they share in the same entry. In a termbase it also means that all languages have equal status.

7.1.3 Term autonomy

All terms are created equal and can be described with the same degree of detail (that is, using all the same fields in the system). All terms that denote a concept are managed as autonomous and repeatable blocks of data categories within a terminological entry.

7.1.4 Data elementarity

All information is separated into individual data categories so that there is only one type of information per data field. For example, it would be a violation of the data structure to include explanatory information in parenthesis after the term in the term field, or to provide both context and definition in the definition field.

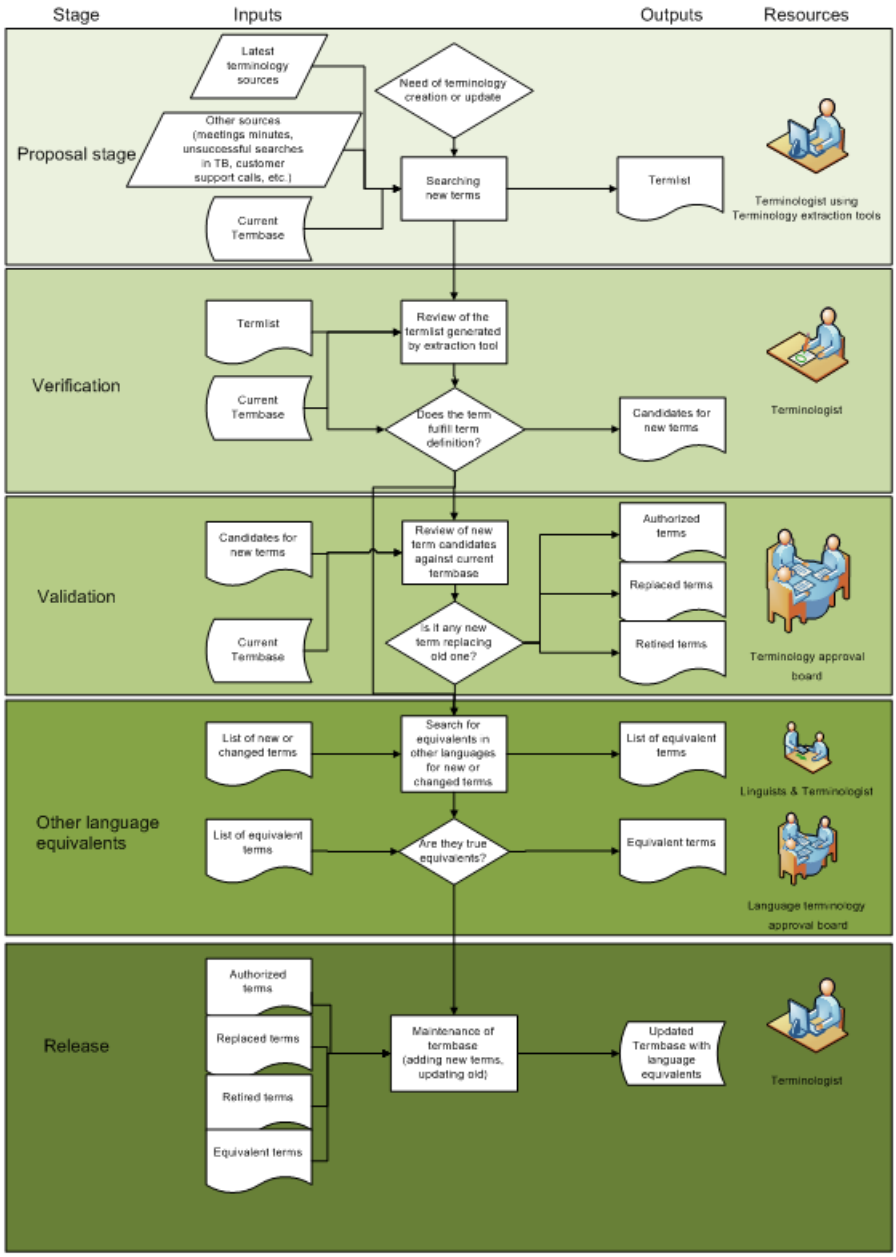
7.2 Process

The process of creating a new term entry commonly includes these steps:

- Proposal
- Verification
- Validation
- Release
- Addition of equivalents in other languages

A term life cycle includes the creation and validation of new entries, rejection of invalid terms, maintenance of existing entries, and mechanisms for retiring terms as they become obsolete. Terms that are incorrect, invalid or obsolete are not removed from the termbase; rather, they are managed and identified by using data categories that indicate their status as deprecated or no longer used.

The following illustration outlines the common steps in a terminology workflow:



7.3 Quality assurance

Quality assurance is needed on three levels:

- For individual terms it means verifying that new terms (neologisms) follow any existing brand strategy, copyright regulations, usability guidelines, and sound term formation.

- For terminological entries it means verifying that information in the each terminological entry is entered accurately, correctly and according to adopted standards. Follow the data management principles of concept orientation, term autonomy and data elementarity.
- Governance mechanisms are required to reconcile terms between groups within the organization.

7.4 Users

To support both the usability and quality of a termbase, the TMS must support access by multiple users, as well as safeguarding the information that it contains.

- The termbase should be available to everyone in the organization and allow concurrent users. Different users may have different types of access (read-only, read-write) based on their roles and needs.
- Multiple concurrent users demands a security model that distinguishes and restricts access rights based on a combination of criteria:
  - Data level: users are granted privileges to either the whole entry or only a portion of the entry
  - Language: users are granted privileges on a language basis
  - Roles: users are granted privileges according to their role (e.g. term consumer, term submitter, term reviewer, term approver)
  - Subject area: users are granted privileges according to their areas of expertise
  - Organizational structure: users are granted privileges according to their position in the organizational structure (e.g. internal vs. external users).

7.5 Reporting

Three common types of reporting needs for terminology management are:

- Status reports typically cover quantity of terms and entries, including information on number of new, changed, and deprecated entries (globally, per language, or per project).
- Maintenance reports are used to detect quality issues and deprecated or obsolete entries in the term database. Quality issues may include invalid data category combinations (for example a definition written for a noun, but part of speech set to ‘verb’), empty fields, and spelling and grammar errors (in text fields).
- Terminology change reports provide information on terminology changes and updates. These can be used to check other resources (e.g. documentation, translation memories) that may be affected by global terminology changes. Change reports are also useful for informing affected teams in the organization of any changes to key terms.

7.6 System integration

To produce terminology and distribute it within your organization, you will need to automate repetitive tasks and termbase-wide operations. You must also ensure that your tools are compatible with each other. The TMS should have a documented API and it should support file formats that are compatible with other tools in areas such as:

- Product development
- Document authoring and knowledge management
- Quality assurance
- Translation
- Term extraction or harvesting
- Customer management and care

7.7 Terminology management tools

Terminology management systems can be seen and classified from many perspectives focusing on one or more different parameters or criteria. When acquiring a TMS, there are four categories of solutions to consider.

7.7.1 Re-purposing existing technologies and software

Description: Spreadsheets and word processing applications are often used to collect lists of important terms. Columns in a table may be used to provide initial classification of data.

Main advantage: Easily available.

Main disadvantage: Not able to do proper, concept-based terminology management; not integrated into translation tools; does not easily accommodate concurrent user or access rights.

7.7.2 Re-purposing existing technologies and software

Description: Spreadsheets and word processing applications are often used to collect lists of important terms. Columns in a table may be used to provide initial classification of data.

Main advantage: Easily available.

Main disadvantage: Not able to do proper, concept-based terminology management; not integrated into translation tools; does not easily accommodate concurrent user or access rights.

7.7.3 Accompanying module in a bigger solution

Description: Translation tools, authoring tools, enterprise search, targeted ontology, and taxonomy management tools have a requirement to store monolingual or multilingual terminology. Therefore they often include some terminology management functions. The feature coverage and level of sophistication ranges from very simple (flat term list) to very sophisticated (feature-rich, concept-oriented).

Main advantage: No extra cost; integrated into solution.

Main disadvantage: Often not able to provide proper terminology management, as features are driven by the parent solution (translation tools or authoring tools often have a smaller scope than a proper TMS).

7.7.4 Dedicated terminology tool

Description: Full-featured software with the primary purpose of storing, maintaining, and distributing terminology data. It may be server and/or web based, or alternatively just a stand-alone software for a translator or technical writer.

Main advantage: Very appropriate for terminology needs.

Main disadvantage: Often not integrated into translation or authoring solutions, extra cost.

Terminology extraction tools

To help identify information about concepts and the terms representing them, terminology extraction tools can be useful. It is often too time-consuming to manually identify the important terms in an information set. Running a terminology extraction tool on a text or group of texts (corpus) automatically produces a list of words and word combinations that might be terms, called “term candidates.” A person (often a terminologist) will then go through the list to determine which of the words and phrases are true terms. The terms selected by the terminologist will often become the focus of further terminological research, with the goal of recording the term and the information about the concept it represents in a terminology record in a termbase.

8.1 Features of term extraction tools

8.1.1 Term in context

Some terminology extraction tools also allow you to view all occurrences of the term in context. Being able to browse all of the sentences in which that term occurs makes it easier to locate other information in the corpus about the term, such as definitions, synonyms, and examples of usage.

8.1.2 Languages

Some extraction tools only process texts in one language, while others can process bitexts (the pair of a source text and its translation, aligned).

8.1.3 Statistical and linguistic engines

Some tools find the terms by using statistics only (which words or word combinations occur most often), while others integrate linguistic rules to help determine which words or word combinations are likely to be terms (as opposed to function words, phrases or random strings of words).

8.1.4 Other features

Some tools can be set so that they only extract terms that occur more than a certain number of times in the processed text (consider a list of terms that occur at least 4 times versus at least 20 times). Some extract single-word terms, others only multi-word terms, others both. Some terminology extraction tools that can process bilingual texts will not only extract a segment that could be a term, but even search in the translation for a segment that might be its equivalent. Some require a person to validate this pairing, while others can be set to select what is statistically the “most likely” match.

It is best to first decide what languages and features you need, then select your terminology extractor accordingly. You will benefit by finding all occurrences of a term and finding terminological information about a term much more quickly and systematically than you would by doing it manually. Once you start using such a tool, you will find it easier to determine which terms you use most often, and which ones need to be researched, so that this information can be shared in a glossary or termbase with everyone who needs to know it.

Common applications of terminology in other tools

9.1 Enforcing terminology when translating

Translators benefit from approved terminology. Therefore, the translation tool used should be able to send the sentence currently being translated to the terminology tool so that known terms are automatically identified and highlighted, and the translators can then choose the approved target terms.

Furthermore, a translation tool may consult a termbase and compare the terms actually used in the translated text with the ones that are stored in the termbase. If a translator has used terminology other than that which is prescribed in the database, the system raises an alert that can be investigated. This terminology “checking” function is similar to a spell checker.

9.2 Checking terminology when writing

Authoring tools can look up terms in the termbase to verify whether the right terms are used. If the termbase contains not only the “approved” terms but has also the undesired “deprecated” terms classified as such, the authoring tool can evaluate this information, allowing it to warn the technical writer and pinpoint terms that should not be used.

9.3 Modelling taxonomies and ontologies

Many organizations develop taxonomies or ontologies to structure their knowledge, product lines, and so forth. While a taxonomy or ontology models and stores the relationship between concepts, terms form the navigable “handles” for browsing the data. Therefore, taxonomies and ontologies store terms in a very structured manner. The terms that are used as labels in taxonomies and ontologies should be standardized within the organization. There is therefore a strong relationship between ontologies, taxonomies and termbases, and their development should be synchronized.

Basic steps for setting up a corporate TMS

Setting up a terminology management system (TMS) and process can be a complex project. Here are some basic steps:

1. Decide what problem you are trying to solve: Increased efficiency (saving duplication of term research effort so localization managers can manage the project and do not have to manage last-minute terminology crises ad hoc.)? Improving communication internally/externally? Cost savings?
2. Identify what resources you already have: Persons willing to be terminology resources for the corporation available computer hardware, servers, networks, etc.
3. Get the terminology expertise you need: Hire or train someone who has a language background (preferably a terminology background) to be the main resource for terminology for your corporation. Make sure everyone knows who he or she is and what they can ask that person.
4. Identify the information that you require for each term: Terms, preferred terms, deprecated terms? For which languages? Author info and date stamping to track updates to records? Definitions? Other information? Make sure your database will have the fields you need.
5. Decide which criteria you'll look at to make your choice. Online web-based terminology database, or a client-server application? Central control, or open like a wiki?
6. Select your tool: Look at the terminology software available and find the one that meets your needs. Some questions you might ask include:

   • How many people might need to connect simultaneously?

   • How many records can it handle?

   • What views does it offer (dictionary style vs. single-concept)?

   • Can you control read/write privileges?

   • How much does it cost and how much can you afford?

   • If you deal with many languages, does the tool treat all languages equally?

   • Which technology suppliers are you already dealing with? Can they offer any solutions? What do they have?

   • Are import and export formats compatible with the various tools the TMS will need to integrate with?

7. Create an implementation plan: Make a plan to implement the tool that includes a training plan for users. Include a communication plan to promote the new process and tools, and to explain why terminology management is important.
8. Define a maintenance plan: Have an ongoing plan to review and update the content of the database as new terms are added and current terms become obsolete. Remind users that the terminology service is there to help them and that they are expected and encouraged to use it.

Best practices for setting up a corporate termbase

Once a terminology management system has been selected, its implementation and maintenance is an on-going cyclical activity. In order to be successful, a TMS implementation must continually engage and educate users as well as improve and increase its functionality.

11.1 Infrastructure

The TMS and termbase will go through several stages of maturity:

- Acquire or build the TMS
- Troubleshoot and/or run a pilot project
- Continuously adapt and customize the system to the needs of the organization

11.2 People

An important aspect of the successful growth and application of a termbase is to engage as many people in the organization as possible. To ensure success, ensure that you:

- Build awareness across the organization. In organizations where terminology management is relatively unknown, it is recommended that this phase of the implementation starts early in order to raise awareness and develop the necessary technical skills.
- Identify key stakeholders and recruit champions (stakeholder representatives who are willing to support the endeavour).
- Develop training modules that are customized to the different user types and scenarios.
- Reward active participation: For example, on a quarterly basis, select a terminology help request and calculate how much time and/or money one terminology search has saved the company (include multipliers such as number of language versions and number of text formats). Recognize the employee's efforts and savings to the company. Reward them if possible, and make that effort known to all employees. Show your employees how such a seemingly small contribution (like asking which synonym is to be preferred and used) saved the company money, and that their contribution is valued.

11.3 Roll-out

To generate the required interest and participation, ensure that terminology management tools and processes are rolled -out to all users and stakeholders.

- Make announcements. Create publicity around the TMS in internal communication channels and events.
- Put the TMS into production.

11.4 Review

To ensure widespread adoption and use of your termbase and TMS, ensure that the users have a voice with regards to changes and improvements:

- Collect user feedback and requirements.
- Identify improvements to processes.
- Develop new functionalities or customizations.

Common risk factors

There are a number of risk factors that need to be accounted for during the early stages of implementing a TMS. These include, but are not necessarily limited to:

12.1 Lack of sponsorship from executives

Management buy-in and understanding of the importance of terminology management to the company's overall communication strategy is critical. Without a committed sponsorship from key executives, it is nearly impossible to realize the full potential of terminology management.

12.2 Lack of buy-in from stakeholders

Terminology management is a collaborative and collective effort. In addition to top-down backing from upper management, bottom-up support from people on the front lines is crucial. Terminology management is truly successful only when each person in the organization either helps feed the database or consults it when writing. A risk factor is tied to executive sponsorship in that a continuous support and participation from stakeholders is only possible when it is made part of their formal objectives and performance measurements, or when the behavior becomes ingrained in the corporate culture.

12.3 Disconnect from product development cycle

Terminology management needs to be an integral part of the product development cycle, from inception to sale, not as an afterthought in the documentation or translation departments.

12.4 Competing projects

Identifying other projects in the organization that may compete for resources (such as budget or personnel) at an early stage provides an opportunity to build on existing efforts, and to avoid duplicate work in terms of time, tools investments, and resources.

12.5 Insufficient resources

Without sufficient resources for staffing and tools, a terminology project can stagnate and fail to produce the desired return on investment.

Corporate language

Companies want to be identified quickly, clearly, and positively by potential clients and the public. Marketing measures designed to shape and strengthen the company's identity may be summarised in a corporate identity or marketing manual. The manual defines, among others, the corporate logos, colours, fonts, titles as well as other graphical elements, and their binding use in official documents.

An equally important element of corporate identity is corporate language, also referred to as corporate communications. A typical corporate language instrument is the corporate style guide, a collection of linguistic rules, for example on how to separate thousands and decimals in numbers, how to write currencies, which language variant to use (if there are regional differences), how to quote, etc. Another pillar of corporate language is corporate terminology; it covers the names of functions, titles and organizational units, product names, internal abbreviations, project names etc.

Corporate terminology is normally managed and published in a termbase (or terminology database) which is used by all employees and suppliers, including translation services, communications, public relations and marketing. Termbases are primarily used in connection with translation memories and machine translation tools, but they can also be combined with other knowledge management systems such as management information systems, document management systems or content management systems, controlled authoring software, and so forth.

People

14.1 Overview of the roles of team members

The concept of terminology management entails considerably more than terminology management software and the integration of systems. It's the "people component" that drives the goals and processes. Furthermore, the identification of the right people who have the right skills who must be in place at the right time is required in order to achieve success. This section presents recommendations for staffing a terminology team.

The core members of a terminology team are the following:

- Executive sponsor
- Corporate (or organizational) steering committee
- Project manager
- Terminologists
- Technical writers and editors
- Task-focused technical teams
- Linguist
- Globalization and localization expert

The wants and needs of various types of end users are gathered in advance of the TMS design process and at various points during the development and testing phases. The TMS enables end users (both internal corporate consumers and clients or external customers) to do their jobs knowledgeably and efficiently.

14.2 Executive sponsor

An executive sponsor is required to ensure that the project has high-level support. The executive sponsor is attuned to the goals and dynamics of the corporation. The executive has relationships with the key stakeholders who are involved in setting strategic goals. Also, the executive understands the challenges, constraints, and the economic environment of the corporation.

With this corporate knowledge, the executive knows how to manage such an innovative initiative at the corporate level. Using knowledge of other corporate plans and priorities, the executive can advise the

team on the general timeline for the project. Cost-benefit justification for terminology management can be very difficult. An executive sponsor can “open doors” at various corporate levels in order to promote acceptance of the project.

14.3 Steering committee

The steering committee is a cross-disciplinary group that sets the vision and strategy for implementing a TMS within the corporation. For example, in a software company, the steering committee might consist of one relatively senior-level person from each of these divisions:

- Globalization and Localization
- Publications
- Research and Development
- Marketing
- Quality Assurance
- Technical Support
- Information Management Systems

Representatives from the key product lines may also be required. This team meets regularly, monthly or bi-monthly, to drive the terminology deployment project to completion.

14.4 Project manager

Terminology management is most often implemented in a matrix environment. That is, a dedicated, full-time Terminology Management group is not always available to do the work. Instead, people are “on loan” from other groups to perform terminology management units of work. Their involvement depends on the particular task. For example, in a matrix environment, it is possible to recruit a technical person to help with implementing a Web service or a writer to test certain aspects of the usability of the tool and its integration into their daily work.

A good project manager is the key to the success of the TMS. Ideally, the project manager understands both the corporate climate and the technology. The project manager is also committed to success, has exceptional communication skills, and has management experience. The project manager is ultimately responsible for synthesizing and directing all aspects of the terminology management initiative. The project manager should have a reporting relationship to the executive sponsor in order to ensure direct communication, as needed.

14.5 Terminologist

Terminologists are personnel whose primary responsibility is to develop and maintain the corporate language assets for two primary target audiences:

- External users of terminology (customers who purchase and use the products or services): Most external terminology emerges from products and corresponding documentation and product user interfaces.
- Internal users of terminology (employees who depend on corporate resources for accomplishing their work): Organizations that participate in and support Research and Development activities need to share a common understanding of the terminology that is used to communicate about the product during its stages of development.

Terminologists look for and record new terminology. They also manage existing terminology by updating records to identify terms that have become obsolete, or by cleaning up any supporting data in the termbase as required.

They communicate with a diverse group of corporate subject-matter experts to determine the appropriate terms for concepts; to develop accurate, precise concepts that conform to globalization requirements; and to populate and maintain the official corporate termbase with the appropriate terminological data. Terminologists are the corporate champions for terminology management, and can be contacted by many departments within a company to provide terminology advice and research.

Terminologists promote the importance of terminology management via timely and relevant articles about terminology topics on the corporate intranet.

While many terminologists work as described above, i.e., as language professionals specialised in the scientific study of the concepts and terms used in a given subject field, there are also numerous examples of terminologists as generalists in the sense that their role is one of facilitation. They facilitate communication by providing tools and resources that allow people to communicate and speak with one voice. They liaise with subject matter experts to help develop net-based, distributed and cooperative terminology processes. They help design and implement terminology management systems and they train teams and individuals to build and maintain term resources.

14.6 Technical writer

Implicit in the technical writer's job is the documentation of concepts that relate to the products that they document. Terms are collected, researched, analysed, and delivered in the document's glossary. The writers with their subject-matter experts in Research and Development are responsible for identifying key terms and drafting definitions. Writers and terminologists work together to populate the termbase with terms and definitions that must undergo a review and approval process. With approved terms and the appropriate export functionality, writers can generate glossaries for inclusion in their documents.

14.7 Technical editor

Technical editors work directly with writers and terminologists to validate terms and definitions according to established grammar and style guidelines. Technical editors often have a broader view of an organization's overall content and content strategy which allows them to identify terminology topics and challenges across typical project or product boundaries. The person given the responsibility for approving terms (this can be a terminologist or editor, for example) approves final glossaries for inclusion in a document and has authorization for final approval of terms and definitions in the official corporate termbase.

14.8 Linguist or computational linguist

A linguist who is skilled in the interpretation and application of languages plays an invaluable role by assisting with data structure design and addressing terminological issues. The linguist might also be responsible for supporting software that can ensure quality for documentation and collateral. Related expertise is helpful when terminology is used to develop taxonomies and ontologies.

14.9 Globalization and localization expert

A globalization and localization expert is also a critical member of the team who helps ensure correct and consistent source text that is culturally neutral and world ready. Requirements and processes that are associated with globalization and localization activities must also be considered and factored into the overall TMS process. This is important regardless of whether terminology is being translated into multiple languages. It is not uncommon for organizations to start out with a focus and a single language and market, only to expand later. It is both easier and less costly to accommodate globalization and localization considerations in TMS system and processes from the start, rather than make such adaptations later.

14.10 Technical teams

Teams of individuals with vision and expertise are essential for implementing a successful TMS. Teams can be formed for various tasks and steps in the process of building and supporting an enterprise terminology management program. Some individuals will serve on a number of teams and in a variety of roles.

- A team is needed to carefully identify the requirements for the corporate TMS and to select who is best able to fulfil those requirements (either an external vendor or an internal development team).
- A team is needed to create a data structure that adequately stores and renders the terminological data appropriately in the TMS.

- A technical team with a skilled system administrator and a database manager are needed to implement the data structure in the TMS and to load the database with valid terminological data.
- Other technical experts are needed to modify existing tools and systems that directly integrate with the TMS. For example, the TMS must work seamlessly with the document authoring system and attendant proprietary and third-party tools that support document production.
- Another team may be needed to create an access point for the TMS and its collateral on the corporate intranet.
- If migrating between TMSs, a team must be in place to cleanse the terminological data in preparation for an export to the new system
- After the TMS has been set up, a testing team or a group of volunteers verifies the exported terminological data, validates the correct implementation of the data structure, and demonstrates that the new TMS is operational.
- A training team is also needed to develop instructional modules for new users of the TMS.
- It is also helpful to have a communications liaison that can help to promote important corporate TMS initiatives.
- Another team devises the process workflow for drafting, reviewing, approving, and accessing terms and definitions in the termbase

Conclusion

15.1 Why manage terminology?

Managing terminology can positively impact profitability, customer relations and the overall image of an organization.

If terminology is not managed, your staff and external suppliers will decide for themselves what terminology to use, and this could be contrary to the image you would like to portray as an organization. Inconsistent terminology can be deemed at best unprofessional, at worst it can be dangerous as it could lead to misuse of products causing bodily harm. In addition, inconsistent terminology within a company adds costs in several ways.

Organizations spend money on marketing and public relations exercises – which can be wasted if terminology in other means of communication delivers a contradictory or inaccurate message. Without strategic management of terminology, employees make variant choices for terms to express a concept, and the company intranet can become unreliable when terminology searches unearth multiple definitions for a term. Unclear direction regarding terminology often necessitates re-work during authoring and translation projects. When incorrect terminology is propagated into products and collateral, the costs of imposing consistency is multiplied.

There can even be misunderstandings regarding the basic requirements of a project due to different use of terminology between departments, so re-work may actually require production tasks to be redone, often at great cost. Inconsistent terminology has an impact in all areas of communication, for better or worse. Thus, proactively establishing consistent terminology across the organization is not just an exercise in establishing good quality writing, but it can improve the effectiveness and efficiency of a business as a whole.

15.2 Who needs to be involved?

Overall, having the right people involved in the design, development, and implementation of a terminology management programme and its associated TMS makes the difference in how well terminology is accepted and used in the organization. The right mix of software can be in place, but a successful deployment depends on leveraging the talent and expertise of the core team as well as the extended organization.

Ideally terminology is standardized as part of product development, so that it can be used consistently in all authoring and packaging, and so that standard translation equivalents can be determined. This moves the focus of terminology away from end-product glossary created by an isolated team; rather, it is a case of key people from across the organization working together to prescribe its own consistent language. The aim is to form an identity, to establish clear communication, and to thereby realize savings in time, resources, and money. As we have mentioned when examining the role of terminology management, successful terminology management fits in and benefits an organization if all stakeholders understand and promote the importance of accuracy and consistency within its communication efforts.

15.3 What is the most effective approach?

Terms should be handled as individual concepts, because the problem of homographs will otherwise cause confusion and errors. Homographs are numerous, and especially problematic for translation (whether by machine or by humans). For example: is “copy" the duplicate of an item, or a piece of text provided for marketing materials? We use the same word in English for two distinct concepts – but the translation equivalent in other languages is likely to be different for each concept. The Terminology Management System stores terms and makes them available for reference by authors and translators so that they can be consistent with prescribed terminology . The chosen TMS should be compatible, and ideally integrated, with the existing authoring and translation systems.

Note that this is not a do-only-once activity. Terminology management needs to be part of the company strategy. Language changes – so does the way we talk about our products and the way we interact with our customers. Such changes need to be managed – accepted and documented if deemed appropriate, or if not appropriate, individuals need to be educated as to what the company requires. Terminology management adds value to all communication within an organization and supports all levels of product and services life cycle and any accompanying communication.

Resources

Please note that this list may not be exhaustive. The field of terminology management is rapidly evolving and new research and papers are continuously being published.

16.1 Standards

Basic principles

- ISO 704 - Terminology Work - Principles and Methods
- ISO 1087-1 - Vocabulary

Data modelling

- ISO 30042 - Systems to manage terminology, knowledge, and content - TermBase eXchange (TBX).

  Can be freely downloaded from: http://www.ttt.org/oscarstandards/
- ISOCat - Data Category Registry (DCR). http://www.isocat.org/
- ISO 16642 - Computer applications in terminology - Terminological Markup Framework (TMF). "http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=32347"
- ISO 26162 - Systems to manage terminology, knowledge and content -- Design, implementation and maintenance of terminology management systems. http://www.iso.org/iso/catalogue_detail.htm?csnumber=43427

16.2 Books and articles

Bernth, A., McCord, M., & Warburton, K. (2003). Terminology extraction for global content management. Terminology, 9(1), 51-69.

Cabré Castellví, M. T. (2003) Theories of terminology: their description, prescription and explanation, Terminology, 9(2), 163-199.

Dubuc, R. (1997). Terminology: A Practical Approach.

Dunne, K. J. (2007). Terminology: ignore it at your peril. Multilingual, April-May 2007 (" http://www.multilingual.com/issueDetail.php?issue=87")

Fidura, Ch. (2007). The benefits of managing terminology with tools. Multilingual, April-May 2007 (" http://www.multilingual.com/issueDetail.php?issue=87")

Childress, M. D. (2007). Terminology work saves more than it costs. Multilingual, April-May 2007 (http://www.multilingual.com/issueDetail.php?issue=87)

Karsch, B. I. (2006). Terminology workflow in the localization process [Electronic version]. In K. Dunne (Ed.), Perspectives on Localization.

Lombard, R. (2006). A practical case for managing source-language terminology [Electronic version]. In K. Dunne (Ed.), Perspectives on Localization.

Massion, F. (2007). Terminology management: a luxury or a necessity? Multilingual, April-May 2007 (http://www.multilingual.com/issueDetail.php?issue=87)

O'Neil, B. (2005). Launching a Corporate Glossary. B-eye: Business Intelligence Network, Article 1014. Retrieved September 20, 2006 from http://www.b-eye-network.com/view/1014.

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

Schmitz, K-D. (2002). Towards a uniform environment for representing terminologies within ISO. Presentation at TKE 2002 in Nancy, France. Retrieved September 20, 2006 from http://tke2002.loria.fr/Doc/workshops/ws2/ws2_kds.ppt.

Schmitz, K-D. (2005). Terminology Data Modeling for Software Localization. Presentation at TKE 2005 in Copenhagen, Denmark. Retrieved September 20, 2006 from http://www.id.cbs.dk/~het/TKE2005/Schmitz.ppt

Straub, D., & Schmitz, K-D. (2010). Successful terminology management in companies. (ISBN:978-3-9812683-1-7) TEKOM study:" http://www.tekom.de/index_neu.jsp?url=/servlet/ControllerGUI?action=voll&id=3162" .

Terminology and Standardization Directorate, Pavel S. & D. Nolet, Translation Bureau. (2001). Handbook of Terminology. Retrieved November 1, 2006 from http://www.translationbureau.gc.ca/pwgsc_internet/fr/publications/gratuit_free/man_termino_e.htm. Gatineau, Quebec: Public Works and Government Services Canada.

Wittner, J. (2007). Unexpected ROI from terminology. Multilingual, April-May 2007 (http://www.multilingual.com/issueDetail.php?issue=87)

Wright, S. E. and G. Budin (Eds.), Handbook of Terminology Management, Vol 1. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wright, S. E. and G. Budin (Eds.), Handbook of Terminology Management: Application-Oriented Terminology Management, Vol 2. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wright, S. E. (2005). ISO TC 37 Standards--Basic Principles of Terminology. Presentation at the Joint Conference for Digital Libraries (JCDL) 2005. Retrieved 20 September, 2006 from http://nkos.slis.kent.edu/2005workshop/TC37.ppt.

Zielinski, D. & Ramírez Safar, Y. (2005). Research meets practice: t-survey 2005. An online survey on terminology extraction and terminology management. Retrieved September 20, 2006, from Linguistic Data Processing Section, Saarland University: http://fr46.uni-saarland.de/download/publs/sdv/t-survey_aslib2005_zielinski.htm.

16.3 Web resources

- TermNet
  http://www.termnet.org
- IATE (InterActive Terminology for Europe)
  http://iate.europa.eu
- Eurotermbank
  http://www.eurotermbank.com
- Termium
  http://www.termiumplus.gc.ca
- Banque de terminologie du Quebec
  http://www.granddictionnaire.com
- TermCat
  http://www.termcat.cat
- United Nations FAO termbase
  http://www.fao.org/faoterm/search/start.do;jsessionid=8B3FA262DA3D26D7908F292E331001CC
- Microsoft Language Search Tool
  http://www.microsoft.com/language/en/us/search.mspx
- IBM Terminology
  http://www-01.ibm.com/software/globalization/terminology/index.jsp

16.4 Online tutorial

The Pavel Terminology Tutorial, Translation Bureau of Canada

http://www.btb.termiumplus.gc.ca/didacticiel-tutorial/lecon-lesson-1/index-eng.html

Index