TSH 35/36 - 2008

Terminology Standardization and Harmonization

ISO/TC 37 "Terminology and other language and content resources"
http://www.iso.org/tc37

Contents/Sommaire   Page

© Infoterm 2008

Offenlegung nach § 24 Mediengesetz:

Terminology Standardization and Harmonization (TSH) ist ein vierteljährlich erscheinendes informationsblatt des Sekretariats des Technischen Komitees ISO/TC 37 "Terminology and other language and content resources" der Internationalen Normungsorganisation (ISO) und des Internationalen Informationszentrums für Terminologie (Infoterm). TSH enthält Informationen und Nachrichten über Ereignisse, Tätigkeiten und Projekte aus dem Bereich der Terminologienormung auf nationaler, regionaler und internationaler Ebene und verfolgt dabei keine parteilichen oder ideologi-schen Zielsetzungen. Ziel dieser Publikation ist es, alle terminologisch tätigen und interessierten Organisationen und Personen über die laufenden Aktivitäten auf dem Gebiet der Terminologienormung zu informieren, aktuelle Informationen und Hilfestellung für ihre berufliche Tätigkeit zu liefern sowie ihre Zusammenarbeit zu fördern.

TSH is a joint publication of the Secretariat of ISO/TC 37 and Infoterm. It has been created in 1989 with the objective to foster communication and cooperation among organizations and individuals involved in terminology standardization and harmonization. It provides information on terminology standardization, especially within the framework of technical Committees, as well as on the results of their activities.

TSH est publié conjointement par le Secrétariat de l'ISO/TC 37 et Infoterm. TSH fut fondé en 1989 afin de stimuler et d'encourager la communication et la coopération entre les organismes et les personnes engagés dans le domaine de la normalisation de la terminologie. Il renseigne sur les activités de normalisation de la terminologie au niveau international ainsi que sur celles au sein des comités techniques.

PREFACE

The Infoterm publication strategy was initiated in 1980 with the support of UNESCO and the assistance of the Translation Bureau of the Canadian government. The publication scheme started off with the Infoterm Newsletter and was extended to TermNet News (TNN, 1980), BIblioTerm (BIT, 1983), STandardTerm (STT, 1989) and Terminology Standardization and Harmonization (TSH, 1989).

Together with STandardTerm, TSH had been conceived as a contribution by Infoterm to ISO/TC 37 and to ISO at large. Till the end of the 1990s, they have been distributed to member bodies and liaisons of ISO/TC 37, all member organizations of ISO, Infoterm, and TermNet, as well as to individual members of ISO/TC 37, IITF (International Institute for Terminology Research) and GTW (Association for Terminology and Knowledge Transfer), in order to inform about activities in the field of terminology standardization.

At the end of the 1990s, STT and TSH were turned into electronic publications – not least due to the financial burden involved in the distribution. Given the fact that
-    the Internet has revolutionized editing, publication and distribution processes,
-    the ISO Concept DataBase (ISO/CDB) is about to become operational soon,
Infoterm at the end of 2008 had to adapt its strategy with respect to periodical publications and the dissemination of information on terminology standardization. Thus, TSH N° 35-36 is the last issue of this series. .

In this connection, we would like to thank our cooperation partners – in particular Infoterm Members and ISO Central Secretariat – for their assistance in publishing TSH on a regular basis over the years. We are confident that there will be more efficient mechanisms to provide similar and probably increased and even but better targeted information via web-based distribution methods and channels soon.

The Editorial Team

NEWS

World Standards Day 2008

World Standards Day is celebrated each year on **14 October** to pay tribute to the efforts of thousands of experts worldwide who collaborate within IEC, ISO and ITU to develop voluntary International Standards that facilitate trade, spread knowledge and disseminate technological advances.

The theme of World Standards Day 2008's message is **"Intelligent and sustainable buildings".** With a world population that has more than doubled since 1950 and is steadily moving into urban areas – half the world population is expected to be living in an urban environment by the end of 2008 – the building and construction industry has grown into one of the largest industry sectors with immense consequences for all three dimensions of sustainable development – economic, social and environmental.

The World Standards Day message is signed by the leaders of the three principal international standardization organizations: Mr. Jacques Régis, President of the International Electrotechnical Commission (IEC), Mr. Håkan Murby, President of the International organization for Standardization (ISO), and Dr. Hamadoun Touré, Secretary-General of the International Telecommunication Union (ITU).

World Standards Day 2008 Conference Standardization and SMEs: a challenge for Europe; what are the keys to success?

Paris, 21 October 2008

On the occasion of World Standards Day 2008, the French Presidency of the Council of the European Union, in cooperation with the European Commission, organised a one-day conference as a follow-up to the conference held in 2006, which concentrated on SMEs and standardization.

The aim of the Conference was to assess the progress made in the meantime in implementing good practices, to identify other good practices since developed across the EU, and to review remaining hurdles.

Entrepreneurs, organizations representing SMEs, the national standards bodies, administrations and other interested parties met to share their experiences and throw new light on the strategic importance of standardization for SMEs, the challenges ahead and the solutions which work.

For the video presentations in French, see:
http://www.ue2008.fr/PFUE/lang/fr/accueil/PFUE-10_2008/PFUE-21.10.2008/colloque_europeen_sur_la_normalisation_et_les_petites_et_moyennes_entreprises?mode=diaporama

For the text in German of the presentation of Ms Stampfl-Blaha, Deputy-Director of the Austrian Standards Institute, see: Page 7 of CONNEX N° 166, Dec. 2008

Towards an increased Contribution from Standardisation to Innovation in Europe

The Competitiveness Council of 4 December 2006, addressing in its conclusions the subject of innovation, stressed the need to enhance the European standards-setting system, and invited the Commission to put forward proposals for action to be taken by the relevant stakeholders.

On 11 March 2008, the Commission adopted a Communication on "Towards an increased contribution from standardisation to innovation in Europe", responding to the invitation of the Council. The Communication places in focus a greater contribution from standardisation to innovation and competitiveness.

On 25 September 2008, the Council adopted the following Conclusions

Annual Meetings of ISO/TC 37

Moscow, 10-15 August 2008

This year, the meetings of ISO/TC 37 held in Moscow, where they were hosted by Russian ISO/TC 37 member STANDARTINFORM, which is also a member of Infoterm (*see INL 126/127).* The meetings were aptly taking place at the Academy of Labour and Social Relations in the South-Eastern part of Moscow.

Like every year in the month of August, experts involved in standardization of methodology and principles concerning terminology and other language and (structured) content resources, met for a week to bring their ongoing projects a step further and discuss topics that might be taken up in the technical committee in the near future. TC 37 experts are representatives of research institutions, NGOs and companies, delegated to contribute and vote by either their national standards organization or a registered liaison that has a membership status with the TC.

Infoterm has been holding the secretariat of ISO/TC 27 on behalf of the Austrian Standards Institute for many years and will hand over this duty to its twinning secretariat CNIS (China National Institute for Standardization) by the end of 2008.

For further information on ISO/TC 37, see: http://www.infoterm.info/standardization/iso_tc_37.php

During that labour-intensive and fruitful week, a number of exciting future work items have been discussed, such as sign languages, simplified language, IPA or proper names. These new work items will open the work of TC 37 to new communities which have a stake in either one of these.

For further information on the work of ISO/TC 37 or how to get involved contact: infopoint@infoterm.org.

ISO PRESS RELEASE

2008-07-15

ISO publishes book+CD on integrated use of management system standards

ISO has published a combined book and CD giving organizations advice on how to make use of management system standards.

The integrated use of management system standards distils the experience and expertise of an ISO task force comprising 16 members drawn from business organizations representing a wide range of countries and sectors, as well as from standards bodies and academia.

"The book provides a good mixture of theory and practice and will be useful for beginners as well as for the experienced," declares Petra Eckl of MLPC International, France, leader of the ISO task force. "It will form a bridge between the increasing number of ISO management system standards that meet specific concerns of organizations and their stakeholders, and the user organization's own and unique management approach."

Contrary to many other types of standard, management system standards cover multiple aspects, levels and functions of an organization and, therefore, their implementation can have a substantial impact on how an organization operates and manages its business processes. In addition, more and more organizations are applying not only one, but a range of management system standards to satisfy their own needs as well as those of external stakeholders.

The book acknowledges that there is a market need for separate management system standards addressing different aspects, issues or risks that organizations need to manage. It provides guidance on how organizations can apply the different standards in a combined way, integrated with their business processes.

Based on the practical experience of organizations that have successfully made integrated use of management system standards, the book identifies a variety of methodologies, tools and practices.

The book, which is in colour and includes numerous graphics, is structured as follows:

- a description of the main characteristics, parts and functions of the management system of an organization
- what sort of requirements are contained in a management system standard, why these are important and how an organization should apply them
- guidance on how requirements from multiple management system standards can be integrated in a combined way within the existing management system of an organization.

Each chapter includes the following distinct sections:

- *Guiding questions* that helps to focus the reader
- *Overview* summarizing the theory of the topic concerned
- A*pproach* featuring the methodologies and tools applied
- *Jim the Baker* – an imaginary character whose company is used to illustrate the implementation of the methodologies and tools
- C*ases in point* with actual practices and examples from case studies
- *Practice* assists readers in applying the principles and methods provided to their own situation.

The accompanying CD (inside the back cover), contains real-life examples extracted from 15 case studies made in a wide variety of situations. Examples are included from companies in Asia, South America and the Middle East, as well as Europe and North America. Multinational companies are included as well as smaller ones, in different sectors of activity, for profit and not-for-profit, applying different sets of management system standards.

For further information, see: http://www.iso.org/iso/pressrelease.htm?refid=Ref1144

ARTICLES FROM ISO FOCUS
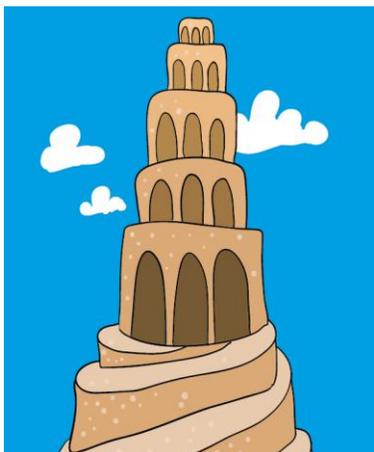
A multicultural effort for a multilingual society

by Xinli Yu, twinned secretariat on behalf of SAC (China) and Christian Galinski, secretary on behalf of ON (Austria), ISO/TC 37 Terminology and other language and content resources

Terminologies are not only fundamental in every subject field, they are indispensable in every language. This is particularly true when the language community wants to develop its language as a tool for scientific-technical communication and participate in the global multilingual information society.

For many countries – and increasingly also for multinational enterprises – terminology policies or strategies are highly suitable management tools serving an array of purposes. Furthermore, structured (multilingual) content – naturally including terminologies – is increasingly becoming the fuel of the global information society, making ISO/TC 37 standards ever more important, for developed and developing countries alike.

Partnering and twinning

The ISO Technical Management Board's (ISO/TMB) concept of partnering and twinning foresaw an arrangement whereby member bodies could work together to build the capacity of the national standards body of a developing country. As the concept took shape, ISO/TC 37 took up the idea and in 2005 discussed the possibilities of partnering and twinning within the ISO/TC 37 framework. As a result, the secretariat of ISO/TC 37 drafted an agreement to function as a twinned secretariat, which was then proposed to the respective ISO members – Standardization Administration of China (SAC) and Austrian Standards Institute (ON) – for submission to ISO/TMB.

The International Information Centre for Terminology (Infoterm), responsible for the operation of the ISO/TC 37 secretariat on behalf of ON since 1971, and the Chinese standardization authorities had been cooperating since the early 1980s under the auspices of UNESCO. The cooperation was intensified in the 1990s, when Infoterm succeeded in bringing about an inter-

governmental *Agreement for the Cooperation in Science and Technology* between The People's Republic of China and Austria. Among others, projects were carried out to support terminology standardization and the development of methodology standards in ISO/TC 37 at international, regional and national levels.

Given this long and fruitful cooperation, the ISO/TMB initiative offered a new way to intensify the cooperation between the China National Institute for Standardization (CNIS) on behalf of SAC, the member body representing China in ISO, and ISO/TC 37 for the benefit of the terminology community worldwide. With the approval of both SAC and ON, respectively the supervising authority of CNIS and the official host of the ISO/TC 37 secretariat, the draft agreement passed and was approved by ISO/TMB in 2005.

Visible kick-off

The twinning agreement immediately bore positive results, putting SAC/CNIS in a position to substantially contribute to the great success of the annual meetings of ISO/TC 37 and the 3rd International Conference on *Terminology, Standardization and Technology Transfer* (TSTT2006). The two events, both taking place in Beijing in August 2006, can be regarded as the visible kick-off of the cooperation. But their collaboration does not stop there.

Even before that, the two partners were already closely cooperating to identify new trends in pertinent research and development, as well as in society and the economy, to be adopted for standardization. Both have been working together towards greater visibility and awareness for ISO/TC 37 standards. Regarding administrative cooperation, a total of three new work items have been identified and taken up by SAC/CNIS, with the logistical and administrative assistance of its twinning partner.

> "Seen from the perspective of both a developing and a developed language, globalization and localization are complementary."

Positive and wider impact

But the twinning cooperation has had a positive impact on ISO as a whole, going beyond ISO/TC 37: CNIS, which also takes care of the secretariat of the East Asia Forum on Terminology (EAFTerm), was one of the key organizers of a large-scale terminology project to record all ISO standardized terminologies in database format. To achieve this tremendous task of recording tens of thousands of terms from all ISO standards published so far, CNIS cooperated with other EAF-Term members in the region, notably Japan and Korea, and the combined expertise from ISO/TC 37.

The project coincided with efforts within ISO to address the issue of the increasing use of databases in standards development. CNIS contributed the terminology data it had extracted from ISO standards to ISO/CS and thereby supported ISO's initiatives in the TMB ad hoc group "Standards as databases". The work of this group resulted in the design and development of the ISO Concept DataBase (ISO/ CDB) to which Infoterm and ISO/TC 37 experts actively contributed.

> "The twinning cooperation has had a positive impact on ISO as a whole, going beyond ISO/TC 37."

In the next phases of the cooperation within EAFTerm, it is planned that the project will concentrate on organizing Chinese, Japanese, Korean and Mongolian versions of the ISO terminology data as well as selections of Chinese, Japanese, Korean and Mongolian standardized terminologies which are planned to be submitted to the ISO Central Secretariat for integration into the ISO/ CDB. This approach may provide a model for the future federation of terminology work and terminology databases among ISO member bodies and between ISO member bodies and the ISO Central Secretariat.

In addition to these activities CNIS and Infoterm are working on the development of new methods for Web-based terminology standardization and training strategies for terminology standardization experts, in order to explore innovative ways to disseminate and apply the standards developed by the technical committee.

Successful transition

Based on the success and the experience achieved by the twinning secretariat function, SAC was the prime candidate for the reallocation of the ISO/TC 37 secretariat after ON had decided to relinquish the secretariat of this committee. Under ISO/TMB Resolution 10/2008, the ISO Technical Management Board entrusted SAC with the responsibility of hosting the secretariat of ISO/TC 37 (to be operated by CNIS on behalf of SAC). SAC is fully committed to further supporting the development of standardization activities in the field of terminology and other language and content resources.

About the authors

Dr. Christian Galinski is Director of the International Information Centre for Terminology (Infoterm), an international NGO based in Vienna, Austria, and founded by UNESCO in 1971. He is secretary for ISO/TC 37 on behalf of the Austrian Standards Institute (ON). He holds a number of positions in terminology organizations worldwide.

Ms. YU Xinli is Vice President of the China National Institute of Standardization (CNIS). She is also responsible for the Chinese National Committee for "Terminology and other language and content resources". She has acted as the "twinning secretary" for ISO/TC 37 since 2006, in which year CNIS hosted the annual ISO/TC 37 meetings in Beijing, China, and also organized the international conference TSTT 2006 (Terminology, Standardization and Technology Transfer).
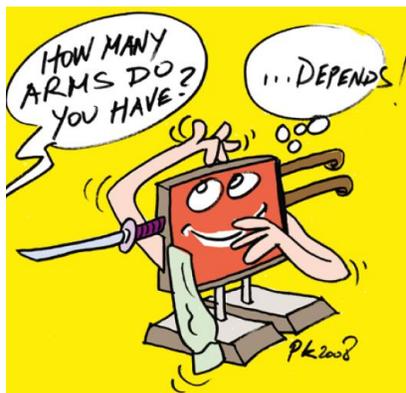
Zooming in on the ISO Concept database

by Reinhard Pohn, co-founder and Managing Director of Paradine, and Reinhard Weissinger, Group Manager, Project Management and Electronic Services, Standards Department, ISO Central Secretariat

Even the smallest error can have enormous consequences. Take for instance the case of the Mars Climate Orbiter, one of the two spacecrafts of the NASA Mars Surveyor 1998 programme. Launched from Cape Canaveral Air Force Station in December 1998, its mission was to study Mars' climate and resources.

The Mars Climate Orbiter was lost on its approach to Mars before beginning its actual mission. Preliminary findings by NASA's Jet Propulsion Laboratory indicate that a failure to recognize and correct an error in a transfer of information between the Mars Climate Orbiter spacecraft team in Colorado and the mission navigation team in California led to the loss of the spacecraft.

Results indicate that one team used Imperial units (e.g. inches, feet and pounds) while the other used metric units for a key spacecraft operation. This information was critical to the manoeuvres required to place the spacecraft in the proper Mars orbit. The mission costs were a total of 327,6 million US dollars [1].

In industry, an error of this type would cause, in addition to huge financial repercussions, substantial delays in time to market. Ensuring that all players are " speaking the same language " is thus a matter of great importance.

What is an arm?

International Standards can help. Their goal is to promote consistency and harmonization, thus ensuring that stakeholders are on the same page. This is particularly important when dealing with global markets. The international nature of these markets offers great opportunities, but also opens the way for a myriad of challenges.

The European Union alone, for example, has over 22 different official languages. How can its countries ensure consistency and harmonization when communicating with each other?

Translation is by no means an easy feat. Take for instance the word " arm ". Out of context, an arm could be a weapon, a part of our body, the arm of a chair, a lever, a sleeve, etc. Clearly, reliable information on the intended meaning and usage is necessary.

International Standards are sources of terminology and requirements, which provide bridges in an increasingly globalizing world. However, to maximize their efficiency they must adapt to the rapid changes in ways of working resulting from developments in information and communication technology.

ISO has thus given some thought to addressing these challenges so that industry can successfully work in a global electronic environment. Its solution takes the form of the ISO Concept database (ISO/CDB), a new state-of-the-art development which plans to release standards in the form of an accessible database.

The traditional and the status quo

Until now, the standardization process has been based mainly on the production of standards in the form of documents. An examination of standards development activities shows that:

- their development takes place within technical committees (TCs), of which the majority are organized vertically based on industry segment and working topic;
- cooperation between committees can be established in the form of liaisons;
- even if one of the basic principles of standardization is to have an industry wide perspective, work on a standard often has a very specific scope;
- many TCs have their own subcommittees (SCs) for terminology;
- TCs and SCs develop and use their own databases to support their work;
- for the most part, standardized code sets like country codes (ISO 3166), language codes (ISO 639) or currency codes (ISO 4217) are published in the form of standards documents;
- in some areas, like graphical symbols, we are seeing the first standards to also be made available in the form of databases;
- national members of ISO have started to offer collections of terminology of national and international standards (e.g. DIN-Term of the German Institute for Standardization).

1)Source: NASA: http://mars.jpl.nasa.gov/msp98/orbiter

However,

- in everyday work, there is a lot of redundancy in investigating and defining terms and definitions ;
- different registration and licensing schemes, inconsistent usage and unclear intellectual property rights statements often confuse potential users of existing databases ;
- the lack of a single approach makes it very hard for both volunteers developing standards and industry using standards to make efficient use of existing databases and avoid misinterpretation.

Getting to the core

However, releasing standards as databases constitutes a Herculean task. The major initial difficulty being the lack of a standard on how to handle this kind of data, as no process had been defined for developing, maintaining and publishing standards data in the form of databases.

To use data in standardization work – or in industry – it must be proven and reliable. Because of a demand for providing such data, some committees had started developing their own solutions, and this had resulted in an increasing and divergent variety of committee databases. Furthermore, most of these databases were, and are, " private " to the committees or organizations involved – they are not designed for wider use by other committees or by industry.

"It will be easy to search for terms, definitions or graphical symbols throughout an entire collection of items."

In this context, the ISO Technical Management Board (TMB) established in 2005 the TMB ad hoc group " Standards as databases ", which comprised representatives from 14 ISO committees as well as from six ISO member bodies.

Its mandate was to review the growing trends in ISO committees to use databases for the maintenance and also, increasingly, for the development of certain types of standardized content which can be referred to as " collections of items ". Examples of such content include terms and definitions, graphical symbols, codes, data sheets and other kinds of structured content, which are either contained in specific types of standards (e.g. vocabulary standards) or constitute part of the content of standards (e.g. a clause containing terms and definitions in addition to other content).

A generic procedure for the development and maintenance of standards in database format was approved by the TMB in 2007[2). The procedure is intended as the basis for an ISO Concept database (ISO/CDB), which contains terms and definitions, graphical symbols, codes and other types of "collections of items". The concept database will comprise content from existing ISO standards, but is also intended to provide a platform for the development of new standards as well as the maintenance of existing standards.

The new concept

After a selection and verification phase, ISO/CS decided to work with Paradine[3) as partner for the ISO Concept database. The project uses Paradine's eptos software – which has been successfully implemented by several standards developing organizations – and integrates it with software applications developed by ISO. The main functions and expectations for the ISO/CDB are:

- to host " concepts ", i.e. items which are already standardized or currently subject to standardization. The ISO/CDB will become the repository for such content in ISO;

2) Annex ST (normative) : *Procedure for the development and maintenance of standards in database format,* of the *ISO Supplement : Procedures specific to ISO*, to the *ISO/IEC Directives,* Part 1: *Procedures for the technical work* (**http://www.iso.org/directives**)
3) Paradine GmbH, Vienna, Austria (**www.paradine.at**)

About the authors

Reinhard Pohn is co-founder and managing director of Paradine GmbH, Vienna, Austria. He has been a member of several national and international standardization committees since 1987. Mr. Pohn received an education in mechanical engineering and marketing, and has 23 years of experience in the software business and mechanical engineering. He has extensive experience in the implementation of metadata dictionaries in standardization bodies, industry associations and corporations.

Reinhard Weissinger is Group Manager for Project Management and Electronic Services in the Standards Department of the ISO Central Secretariat. Prior to this position he worked for several years for the ISO Technical Management Board. Before joining ISO he worked for several years as a long-term expert for the German Society for Technical Cooperation (GTZ) in Beijing in the intergovernmental project *Establishment of a standards information system in the People's Republic of China.*

The language of content-creating communities
by Gerard Meijssen, Wikimedia foundation

If you are looking for an example of " content-creating communities ", you will find many in the Wikimedia Foundation (WMF), which is best known for its English language Wikipedia. Much less known are the other 264 separate language Wikipedias, whose differences in size and quality are profound. In fact, some of these projects are effectively dead – like those in Herero and Tokipona, which not only do not have a community, they also do not even have a single article.
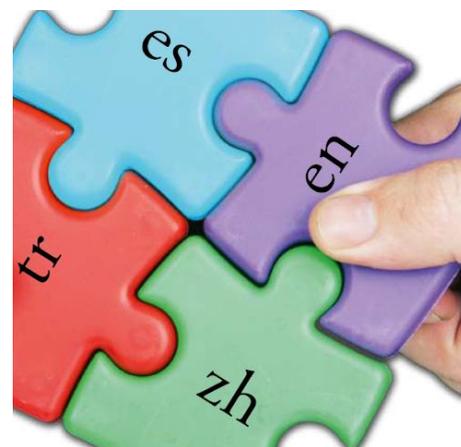
When a request for a Wikipedia in a new language is made, a number of prerequisites must be fulfilled. These requirements ensure that new projects will be viable and benefit the readers of these languages. For example, localization of the most visible messages of the MediaWiki software is necessary to help readers navigate the Wiki. But the first and most important requirement is that the language be indeed a language!

Making life easier

The International Standard ISO 639-3, *Codes for the representation of names of languages* – Part 3 : *Alpha-3 code for comprehensive coverage of languages,* is the most comprehensive list of languages. It covers living, extinct, ancient and constructed languages, whether major or minor, written or unwritten. It already includes over 7000 languages, and the ISO 639-3 registration authority can be requested to consider more linguistic entities as a language, in accordance with the procedures of the standard.

For Wikimedia Foundation's language committee, this makes life a lot easier. If it is not registered in accordance with the standard, you can request that your language be included. It becomes essentially an issue that we do not need to address.

However, this approach is not entirely problem-free. Languages new to ISO 639-3 are not

RFC 4646[1] and consequently, according to the best common practice (BCP), there are no formal codes for publishing content in these languages on the Internet. Another issue is that some of the recognized languages are little more than dialects.

Originally the Wikipedia Web address was simply **http://wikipedia.org**. However, when it expanded into multiple languages, a language indicator was added and the English language version adopted " en " as a prefix. Typically the RFC 4646 provided the language codes, but when a project was requested for a language that did not have a code, it had to be made up, which gave us the **als.wikipedia.org** (German dialects) and the **roa-rup.wikipedia.org** (Aromanian), among several others[2].

Only one Wikipedia is allowed per language. Although logical, this can create issues with some languages. Chinese, for instance, is written in at least two scripts : the Simplified variant and the Traditional variant (registered respectively as Hans and Hant under ISO 15924, *Information and documentation – Codes for the representation of names of scripts*). This was solved by providing a transcription system. For some languages it can be more complicated – for Kurdish, Krim Tatar, Fiji Hindi and Kazakh, for instance, Wikipedia has user interfaces in multiple scripts.

In the end however, people speaking the same language in different locales have to work together to compromise and accept each other's vocabulary and orthography.

This is great from an ideological point of view because it forces people to collaborate, but unfortunately it sometimes does not work this way.

Eating our own dog food

When standards define your approach, it has consequences for downstream products. When the language of a text is defined by its metadata, terminological support systems need to be aware of these languages and provide support. In the *OmegaWiki* project, languages have been accepted based on ISO 639-3.

OmegaWiki aims to bring together information from lexical, terminological and ontological resources in all languages. As a consequence of " eating our own dog food ", we aim to provide information in the language of the user.

Given that *OmegaWiki* stores its data in a database, we are able to utilize the data for multiple applications. We have a vocabulary trainer in production and are developing support for a multilingual categorization system for Wikimedia Foundation's *Commons* project. *Commons* is a project with some three million freely-licensed images. Currently however, all these images are tagged only in English, which leaves almost 50 % of Wikipedia users without a rich image resource accessible in their language.

The problem we are facing now is that ISO 639-3 is concerned only with languages themselves, whilst we also need to be able to recognize linguistic entities like dialects and orthographies.

Technically an orthography can be determined with the use of spell checkers for instance. A spell checker is also a valuable tool to check errors in a document which has gone through an optical character reader (OCR). However, for these to work, the spell checkers must support the orthography of the document involved.

Orthographies and dialects should find their place in ISO 639-6, *Codes for the representation of names of languages* – Part 6 : *Alpha-4 representation for comprehensive coverage of language variants*, currently under development. This next generation iteration of ISO 639 aims to bring together the information of all the ISO 639 standards and provide information on all the linguistic entities in a hierarchical database.

1) An " Internet best current practices" document developed by the Internet Engineering Task Force (IETF), defining tags for identifying languages.
2) People requested new Wikipedias in the belief that once a project existed, editors would come. This worked for some languages, but did not for many others, as can be seen from the List of Wikipedias (see **Box**).

Since *Wikipedia* is just one project among others within the Wikimedia Foundation, you will find that other projects like *Wiktionary* and *Wikisource* will also benefit.

- *Wiktionary*, the " dictionary and stuff " project, has information on different orthographies, such as, for example, British and American English, or Serbian in Latin and Cyrillic characters. Obviously, the BCP allows for all of these, but the problem is that new entries are only added as requested, and associated linguistic entities are not considered. Consequently, the author's preference would be for an approach that does take all the known entities into account and that is extendable.

- *Wikisource* is an online library project where sources in languages are worked on. They can be anything from books, pamphlets and music scores. The works, and therefore the language used, can be really old and dated. A more precise determination of the language would help build a corpus on the Internet in that specific source. As more content becomes available online, any additional support that metadata can offer will help in retrieving the relevant content.

Becoming a way of life

With its 2,5 million articles, the English Wikipedia has become a corpus that includes something about everything. As scientists of many disciplines have observed, Wikipedia is where people invariably go for their initial information. There have been efforts to improve the domain coverage, with the result that specialized Web sites are linking to Wikipedia articles, which themselves often link to related articles in other languages.

Analyses of these links, of the templates and of the text have shown that much information can be gleaned from Wikipedias. This has given rise to other projects, such as *Wikiwords* – a rich terminological resource, *Dbpedia* – a project to create semantic Web content by data mining the English language Wikipedia, and *Wikiprofessional* – providing semantic support for people using the Wiki.

The continuing evolution of community-created content is potentially without limit. As the content grows and more information becomes available under free licence, it will become possible to link into the new information. As translations become available for concepts, it will become possible to find articles in other languages. When Commons is no longer English only, the freely-licensed images that it makes available will be found by school children around the world – members of a generation that will quite possibly grow up accustomed to having access to a wide range of community-created content.

Resources

Wikimedia Foundation: http://wikimediafoundation.org

List of Wikipedias: http://meta.wikimedia.org/wiki/List_of_Wikipedias

Commons: http://commons.wikimedia.org

Dbpedia: http://dbpedia.org

OmegaWiki: http://omegawiki.org

Wikiprofessional: http://www.wikiprofessional.org

Wikisource: http://wikisource.org

Wikiwords: http://www.wikiwords.org

Wiktionary: http://wiktionary.org

Standards and content-creating communities

One issue often discussed with regard to standards is the question of access to users. Many virtual content-creating communities have adopted open/free content practices and the content they provide is freely available to all comers. In the author's opinion, applicable standards, too, should be freely available to further drive their adoption.

International Standards provide formal, structured and standardized information, but their availability at a price can hinder appreciation and adoption of the work involved among some

users. This is particularly relevant for users that do not have the same purchasing power as industry, such as content-creating communities like those within the Wikimedia Foundation.

The virtual community is growing at a tremendous rate, expanding into and originating a diversity of innovative projects that are revolutionizing the way information is stored and communicated. As the world moves into this global era, giving rise to inventive transforming projects such as those of the Wikimedia Foundation, International Standards become increasingly relevant to reliably link global differences. However, like the revolutions that are taking place online, we await the developments that will follow in the international standardization scene, as this community too must adapt to the changes of this cyber age.

About the author

Gerard Meijssen has been involved in "Wikis" and communities for a long time. His involvement in lexical content led to his starting the *OmegaWiki* project. He is a member of the language committee of the Wikimedia Foundation, has spoken at many conferences and advocates the use of standards as a matter of principle.

"The two foregoing articles first appeared in the November 2008 issue of ISO Focus - The Magazine of the International Organization for Standardization - and is reproduced here with the permission of ISO Central Secretariat (www.iso.org). A one-year subscription costs 158 Swiss Francs. Subscription enquiries: sales@iso.org "
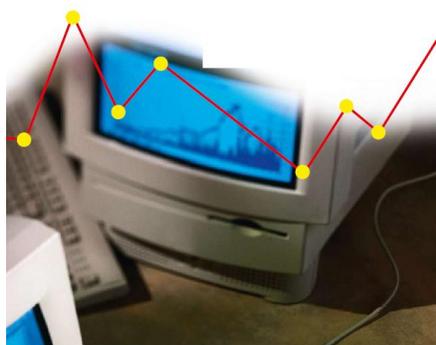
The crucial role of statistical methods

by Christophe Perruchet, Chair of ISO/TC 69, Applications of statistical methods

The production, collection, analysis, presentation and interpretation of data play a key role in industry and service sectors, in scientific research and in human sciences. There is little point in collecting data if it is not going to be analysed and interpreted with a view to enlightening human actions or to further knowledge of phenomena. The aim of statistics should be to transform raw data into usable, understandable and communicable information.

Statistical methods are needed for assessing measurement uncertainty, for calibration, monitoring and improving measurement processes at the producer's site. They are also required by the various agencies involved in testing, verifying conformance and validating the producer's quality and environmental management systems.

In today's competitive industries, and particularly so far as statistical methods are concerned, there is external pressure to demonstrate that good current practice is in place and used, and to have independent agencies certifying that this is really the case. This pressure is creating a growing demand for international standards on statistical methods.

To everyone's advantage

ISO technical committee ISO/TC 69, *Applications of statistical methods,* is responsible for the development of International Standards on the application of statistical methods. Its standards are

used by other ISO technical committees, manufacturers, and the continuously growing industry dedicated to certification, e.g. to ISO 9001:2000 on quality management.

ISO/TC 69 aims to develop and maintain an integrated system of generic standards reflecting good current practice of statistical thinking, which will enable organizations to identify and implement all relevant statistical considerations whenever data are generated, collected, analysed, presented, evaluated and/or interpreted.

"The aim of statistics should be to transform raw data into usable, understandable and communicable information."

A universe of statistical methods

Good practice used by all

The Secretariat of ISO/TC 69 is held by the ISO member for France, AFNOR. An enquiry undertaken in 1996 by ISO/TC 69 showed that:

- more than 400 International Standards of other ISO committees cited standards developed by ISO/TC 69 as normative references;
- each International Standard developed by ISO/TC 69 is cited in at least one standard developed by another ISO committee.

A survey undertaken in 1997 by ISO/TC 69 of its national standardization bodies (14 country respondents) showed:

- 256 national adoptions of the 33 ISO/TC 69 International Standards;
- on average, each country has adopted about 18 ISO/TC 69 standards;
- on average, each standard is adopted by around eight countries.

Key figures

1948 creation of ISO/TC 69

59 country members

29 active subcommittees and groups

73 published ISO standards

18 standards under development

6 proposed new work items

The technical committee is composed of five subcommittees, each of which complements the work of the others.

Terminology and symbols
The terminology standards developed by subcommittee ISO/TC 69/SC 1, *Terminology and symbols*, are aligned with the business environment and can be applied to products and services. The subcommittee serves as both internal and external consultant on terminology matters in this field. To harmonize with other terminology documents, SC 1 ensures that every term and symbol used in the ISO/TC 69 standards has been compiled and annotated to designate preferred terms and symbols. This resource will be especially valuable to technical committees and other experts drafting International Standards and related documents.

Process management
Subcommittee ISO/TC 69/SC 4, *Applications of statistical methods in process management,* has developed statistical methods in the form of control charts, process capability, and process management strategies that are applicable to every enterprise that has continuous improvement as one of its objectives. The economic success of enterprises is enhanced by the application of these tools and techniques, which are essential to any organization in pursuit of lean concepts. In fact, SC 4 has provided ISO/TC 176, *Quality management and quality assurance,* with the statistical tools to enhance the ISO 9001 quality management systems. The same process would be applicable to any standards group.

Acceptance sampling
ISO/TC 69/SC 5, *Acceptance sampling*, is responsible for the ISO 2859 series of standards on inspection by attributes, and the ISO 3951 series on inspection by variables. These useful standards apply just as much to services as to products. The sampling systems permit the level of inspection to be reduced as the level of trust in the producer increases. ISO 2859-1 is widely used in international trade, for example, in the seafood and pharmaceutical industries, among others.

Measurement methods and results
Subcommittee SC 6, *Measurement methods and results,* is responsible for maintenance and development of generic standards and deliverables in such fields as accuracy of measurement methods and results, aspects of the preparation and use of reference materials. SC 6 standards are also generic and apply to both services and products. The ISO 5725 series on accuracy (trueness and precision) of measurement methods and results, is widely used in laboratory testing environments and for conformity assessment. It was decided in 2007 to work on revising the series.

ISO/TS 21748:2004, *Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation,* provides a common approach to measurement uncertainty and relates to the *Guide to the expression of uncertainty in measurement* (GUM). ISO 13528:2005, *Statistical methods for use in proficiency testing by interlaboratory comparisons,* complements both parts of ISO/IEC Guide 43, *Proficiency testing by interlaboratory comparisons.*

Six Sigma

SC 7 is a newly created subcommittee, *Applications of statistical and related techniques for the implementation of Six Sigma*, resulting from an approach made to ISO/TC 69 by the Six Sigma community to provide documents describing statistical procedures that will be useful to those engaged in the practices of Six Sigma. The nature of the task is that the work items will be aligned horizontally across the traditional ISO/TC 69 structure. ISO/TC 69/SC 7 will have the task of developing a framework for these new documents and of managing the work items, with the objective of:

• making the tasks easier to handle within ISO/TC 69 ;
• providing good external visibility of the work being undertaken ; and
• providing consistency in the "look and feel" of the documents that target this new ISO/TC 69 stakeholder community.

The aim is to create documents that provide guidance on the application of statistical procedures for the Six Sigma and related communities, both manufacturing and business types, who strive to improve products and processes. Because these communities are characterized by a focus on speed to obtain results, a basic knowledge of statistics, and a propensity to use software applications to guide them through the mathematical challenges of statistical methods, the documents to be developed will have the objective of satisfying these characteristics.

Perspectives and new fields

To meet the technical requirements of ISO/IEC 17025:2005, it is necessary for testing and calibration laboratories to understand and adopt several statistical methods. Over the years, ISO/TC 69 has developed a wealth of statistical standards that can be used for this purpose.

Six Sigma is a set of practices originally developed by Motorola to systematically improve processes by eliminating defects. A defect is defined as nonconformity of a product or service to its specifications.

While the particulars of the methodology were originally formulated by Bill Smith at Motorola in 1986, Six Sigma drew also on the quality improvement methodologies which had grown out of the six preceding decades, such as quality control and TQM (total quality management). Like these, Six Sigma put emphasis on:

• continuous efforts to reduce variation in process outputs is key to business success;
• manufacturing and business processes can be measured, analysed, improved and controlled;
• succeeding at achieving sustained quality improvement requires commitment from the entire organization, particularly from top-level management.

A document on the application of statistical standards to the clauses of ISO/IEC 17025:2005 is now being developed. This document will give an overview and guidance on the selection of International Standards, guides, technical reports and draft International Standards on statistical methods developed by ISO/TC 69, other ISO technical committees and international organizations. It is presented from the user perspective in the form of a table, comparing the relevant ISO/IEC 17025 sub-clauses against these standards documents.

About the author

Christophe Perruchet has been Chair of ISO/TC 69 since 1994. He joined UTAC (a technical organization providing services to the automotive industry and administrations) in 1984 and is involved in the management and analysis of large databases related to road safety and environment. He is also very active in other ISO committees, such as the ISO technical advisory group TAG 4, *Metrology*, and the ISO committee on conformity assessment (ISO/CASCO) working group WG 28, *Proficiency testing*.