(ISSN1015-4825)TSH31-2007

# Terminology Standardization and Harmonization

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

## Contents / Sommaire



**IMPRESSUM**
**Medieninhaber, Redaktion und Hersteller:**
ISO/TC 37 Sekretariat
c/o Internationales Informationszentrum für Terminologie (Infoterm)
Mariahilfer Strasse 123/3
1060 Wien, ÖSTERREICH
Telefon: +43-664 3446 181
E-mail: infopoint@infoterm.org
URL: http://linux.infoterm.org
**Editor / Rédacteur:** Christian Galinski, Anja Drame
**Layout:** Blanca Nájera



**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.


Commentators talk of the ubiquity of the Web, while discussions elsewhere concern the digital divide and the lack of content in a majority of the world's languages.

Although the predominance of use of the English language in readable Web content is gradually changing, a variety of studies demonstrate that the Web does not present a reliable surrogate for the use of languages in the world. This is possibly because the capability for representing these languages and dialects within these languages has been lacking.

Developments are currently beginning to emerge from ISO/TC 37, Terminology and other language and content resources, in respect of a significant expansion of the well-known series of International Standards, ISO 639, *Codes for the representation of names of languages*.

This is intended to provide a building block of basic identifiers (metadata) with which to index and retrieve the potential
content of a truly diverse and multilingual information society. Previously, standards ISO 639-1 and ISO 639-2 provided for around 400 language identifiers.

Various estimates placed the number of languages in the world between 6 000 and 8 000, and the recently issued ISO 639-3 expands on the existing 400 to produce a set of over 7 500 language identifiers.

Languages are used in different ways, and some languages have a number of different ways in which they can be written, spoken or signed. Identification
of these different ways pushes the total number of identifiers upwards of 30 000, and it becomes clear that until now the ISO 639 series has catered for a very small proportion of the true diversity of "languages ".

The 639 series of ISO standards is generically titled *Codes for the representation of names of languages*, and is expected soon to be divided into parts similar to those shown in the table below.

At its most basic, the expansion from 400 metadata identifiers to upwards of 30 000 would provide support for a global thesaurus of names of languages, in every language of the world − some 400 million potential names, although the true figure is initially likely to be significantly lower.

The developments in the ISO 639 series offer the potential for the reuse of research materials which document the languages, the sum of which − the world's knowledge of the languages of the world − could be significantly greater than its parts.

ISO 639 forms a basic standard for many of the application areas for which ISO/TC 37 develops standards, and within which the ISO/TC 37 standards
are used.

| Title of standard | Status | Registration authority | Number of identifiers (approx.) |
|---|---|---|---|
| ISO 639-1, *Part 1: Alpha-2 code* | Published (2002) | InfoTerm | 150 |
| ISO 639-2, *Part 2: Alpha-3 code* | Published (1998) | Library of Congress (LoC) | 400 |
| ISO 639-3, *Part 3: Alpha-3 code for comprehensive coverage of languages* | Published (2007) | Summer Institute of Linguistics (SIL) | 7 000 |
| ISO 639-4, *Part 4: Implementation guidelines and general principles for language coding* | Expected late 2007 | n/a | n/a |
| ISO 639-5, *Part 5: Alpha-3 code for language families and groups* | Expected late 2007 | TBC | 100 |
| ISO 639-6, *Part 6: Alpha-4 representation for comprehensive coverage of language variation* | Expected early 2008 | GeoLang | 25 000 |

Work is already in progress in the Internet community through the Internet Engineering Task Force (IETF) to make use of these emerging standards, with a newer version of the IETF's language identification, incorporating ISO 639-3, expected shortly.

### ISO/TC 37 Annual meeting

The annual meeting of ISO/TC 37 will be held August 11-18 2007 in Provo, Utah. There will a one-day conference on August 13 on the "Pragmatic Applications of Standards" to look at the wide variety of applications of standards terminology to fields such as information technology, e-commerce and a wide variety of other fields.

For more information on the meeting, please contact: Sue Ellen Wright, **sellenwright@gmail.com**

Historically, use of the IETF's output has been made by the eXtensible Markup Language (XML) – this development will eventually allow for new contributors to make good quality identification within their XML documents.

There is interest, also, in the "Multilingual Internet " – described by some as a major element of the Next Generation Internet – being able to support domain names, e-mail addresses and other types of publicly readable protocol content in character sets other than ASCII.

The need for international country codes has been identified and a New Work Item Proposal submitted to ISO/TC 46/WG 2, *Coding of country names and related entities*. The project, proposed by BSI, aims to establish a joint working group between ISO/TC 37 and ISO/TC 46 and to set up liaisons with interested external organizations.

Further potential exists for these standards to support future generations of current Web-based technologies. For example, a future generation of search engines along the lines of Accoona, that already allows specialization of search by language identifier, might enable searches to be specialized for specific written forms of languages. Future versions of the video sharing website YouTube might allow for searches to accurately index the languages spoken in the video clips, and other so-called Web 2.0 applications could similarly
benefit.

Developments within the ISO 639 series were discussed at the Standards for Global Business Conference held in Vienna 14-15 November 2006, at which developers involved with the ISO 639 series discussed collaboration with the OmegaWiki project, a community-based website for the documentation of information
about languages.

OmegaWiki will support the collection and collation of information about languages by the communities that use them.

This process will assist the verification and validation of language information by the newly-formed World Language Documentation Centre (WLDC), enabling the registration authority of ISO 639 part 6, GeoLang Ltd, to ensure the application of a full verification and validation methodology for the identifiers.

The collaboration has been agreed initially for ISO 639-6, the most ambitious of these standards yet, having to map to the existing parts of ISO 639,take account of existing systems and support the interoperability with such systems.

The World Language Documentation Centre has been formed as an association of world experts and will act as a gatekeeper between language communities and the ISO 639-6 Registry, GeoLang Ltd.

The schedule for publication of ISO 639-6 should enable its availability during 2008, a year proposed as the UN International Year of Languages.

## About the authors

**Dr. Lee Gillam** is a Research Fellow in the Department of Computing at the University of Surrey and Director of GeoLang. Involvement with ISO/TC 37 comes through the British Standards Institution (BSI), efforts supported, in part, by the European Union's eContent programme of research under the Linguistic Infrastructure for Interoperable Resources and Systems (LIRICS). Research interests and publications encompass metadata and ontology, knowledge understanding, and high performance computing. Debbie Garside is Managing Director of GeoLang Ltd and CEO of the World Language Documentation Centre.

**Ms. Garside** is also a member of the Multilingual Internet Names Consortium (MINC) Board and a member of its Secretariat as well as a WikiMedia Foundation Advisory Board Member. She is Convenor of ISO/TC 37/SC 2/WG 1/TG 2, Editor of ISO 639-6 and Chair of BSI mirror committee TS/1/-1. She is a member of the Country Code Names Supporting Organization and Government Advisory Committee for Internationalized Domain Names (ccNSO-GAC IDN) Joint Working Group. Ms. Garside's research interests encompass internationalization, morphology and human genetic linguistics.



# Health information terminologies: standards and practice

*By Christopher G. Chute, MD Dr PH, Professor of Medical Informatics,*
*and Chair of the Division of Biomedical Informatics at Mayo Clinic in Rochester,*
*Maine and Heather Grain, Head, Health Informatics in the School of Public Health*
*at La Trobe University, Victoria, Australia*

## Health information standards

The advent of human genomic applications in healthcare reminds us that rapid advances in health technology are well upon us. Healthcare today is vastly more complex than it was just 20 years ago, relying more fundamentally upon knowledge, information, data, and their interactions. Furthermore, ensuring the ready availability of information in computers for clinical decision making is now a requirement.

The data that confront clinicians and patients poses new challenges for synthesizing and digesting that information. Clearly, the very real requirement that such information be comparable and consistent in its representation and presentation is a made-to-order task for standards development and implementation.

## Semantic interoperability

One of the major challenges of the explosion of information and the capacity to generate, transmit, and share such information, is making sure that the information can “plug and play” in new environments and distant locations. The problems of interoperability, or ensuring that information can be sent by a source machine and read by a target, has been a long-standing challenge in computer science and health informatics.

Traditional interoperability has largely focused on syntactic interchange, which means the formats and layout of data must be the same. However, to be useful in healthcare, information must not only get from one place to another in a consistent format, it must get there in a way that can be understood.

Decision support techniques and patient safety now demand that computers – as well as people – must be able to understand this information. The only solution for what is therefore called “semantic interoperability” is that the contents of messages exchanged for clinical care must conform not only to a standard syntax, but must also contain standard vocabulary elements.

## The central role of terminology

Terminology standards have languished in the background for decades, relegated to mortality classifications or simple drop-down categories of demographics and status measures.

Source: Siemens AG

The emergence of a vocabulary that can exchange detailed clinical information, and be usable in decision support software environments, is a recent innovation. Part of this is due to the technology for authoring and rendering complex biomedical terminologies, attributable in part to innovations in computer science, that make available simplified versions of terminology which are called description logics.

The simplified logics balance the expressive power available to terminology authors with the computational complexity of logical specifications. No computer can manage the full complexityinherent with first-order predicate logic, though many simple variations of description logics do an outstanding job of providing expressive authoring capability while ensuring that the resulting terminologies are computable in reasonable time.

Description logics have fundamentally transformed human capacity to author and consistently maintain complex terminologies. They can support the more complex and challenging use cases that arise in biomedical science, be it in a practice application or in research data interchange.

**“The creation and management of terminologies has become a central challenge in making health information standards work.”**

The creation and management of terminologies has become a central challenge in making health information standards work. The next chapter in the health terminology implementation agenda is to ensure that these terminologies are appropriately broken up into manageable units that are appropriate for a given use case.

This is often called the value-set problem, which means that it is not sufficient to generate complex terminologies, but they must be divided up into useful segments to be plugged into specific applications.

## ISO’s role in health standards

ISO/TC 215, *Health informatics*, was founded in 1998 to begin addressing the complex issues of biomedical data interchange. One of the core working groups formed at the inception of this technical committee, WG 3, focuses on vocabulary. The working group was originally called *Health content representation*, and has since been re-designated *Semantic content*.

The *Semantic content* working group has not attempted to create reusable health terminologies for application throughout the biomedical health arenas, but rather to focus on principles and methods by which such terminologies can be generated.

The relatively small working group does not presume to have the domain depth or knowledge to create all terminologies that might be needed for biomedicine, but it does have substantial experience in how such terminologies can and should be developed. Thus, WG 3 standards and reports specify:

- what constitutes a good terminology;
- a formal glossary of terms and concepts used in creating terminology and semantic content;
- how terminologies might be integrated into a clinical domain space as exemplified by the nursing terminology information model effort;
- policy issues associated with the creation of terminologies intended for broad application throughout bio-health.

One of the core issues associated with terminology use is the intellectual property restrictions that are placed on many health information terminologies. Because terminologies are effectively the "periodic tables" of health information, barriers to their access and use, be they economic or other modes of restricted access, can greatly inhibit the delivery and practice of quality health care throughout the world.

For these reasons, many of the policies and guidelines that are emerging in recommended good terminology practices ensure that fair and reasonable access to terminology content be a core principle of good practice.

In summary, the domain of health information terminology development has emerged from a static backwater to a highly dynamic forefront of health information technology and standards development.
As sophisticated semantic content based on description logic technologies begins to become widely available, we anticipate that the next decade of health information terminology standard development will be even more exciting.

## Selected project summaries

- ISO 17115, Health informatics -- Vocabulary for terminological systems

A formal glossary of term and concepts used in the authoring, maintenance, and application of terminologies.

- ISO/TS 17117, Health informatics -- Controlled health terminology -- Structure and high-level indicators

A project which revises an earlier technical specification on good terminology practices and criteria for evaluation.

- ISO 22789, Health informatics – Conceptual framework for patient findings and problems in terminologies

Guidance on the application of terminologies to the specific problem of clinical findings and observations.

- ISO 18104, Health informatics -- Integration of a reference terminology model for nursing

A sophisticated specification of how a model for describing nursing observations fits within an information model for nursing documentation.

- ISO/HL7 27951, Health informatics -- Common terminology services, release 1

An adoption of an HL7 (Health Level 7) standards document from ANSI about how terminologies can be accessed using a common set of programming interfaces.

# Nanotechnology: the terminology challenge


*by Clive Willis, Convenor, ISO/TC 229/WG 1*

## The need for early standards

Nanotechnology is a relatively new domain that holds major potential gains for both economic performance and improved social programmes. As a result, many nations have afforded a very high priority for basic research activities in nanotechnology and many major corporations are exploring the potential of nanotechnology to improve the performance and quality of their products and production processes.

Although relatively few products are currently in world markets, there is a very significant public interest in ensuring that the introduction of nanotechnology into consumer products and health care services and other sectors is done in a way that does not have a negative impact on health, the environment and the quality of life.

This widely held public position has led national governments to place a high priority on ensuring that appropriate regulatory systems for nanotechnology are in place early and that the standards upon which such systems will be based are reasonably harmonized on a global basis.

It is for these reasons that ISO has chosen to launch early efforts to define a comprehensive set of science-based standards for nanotechnology. National and international organizations, such as the Organisation for Economic Co-operation and Development (OECD), have indicated that

they may defer to the ISO process the task of building a coherent terminology and standards base upon which governance systems can be established.

An important key to the development of nanotechnologies is a coherent and comprehensive vocabulary and terminology for nanotechnology. This is necessary to avoid the vast and confusing set of definitions that has been evolving as individuals and corporations choose their own terms to describe their interests in nanotechnology.

Creating a coherent and consistent terminology is also important for commercial and general communications purposes since agreement on what everyone is talking about (or buying or selling) is critical to rational commercial and public discourse.

## The challenge of integration into existing standards

Since nanotechnology represents a group of technologies that will lead to applications across many, if not all, industrial sectors and social programme areas, the development of coherent standards for nanotechnology is a major challenge.

Concerns are sharply focused on how definitions and terminology for nanotechnology will be integrated into the spectrum of existing standards and how they will be interpreted, in particular in commercial security, health and environmental fields.

The challenge then is to respond to the urgent need to develop a comprehensive standard vocabulary and terminology for the nanotechnology field, while at the same time ensuring that the impact upon other standards development fields is as seamless as possible and without losing the coherent basis upon which clear regulations for nanotechnology can be written or transactions conducted.

## The challenge of a single, basic definition

Nanotechnology is built upon the recognition that the properties of materials exhibit different attributes when dimensions approach the nanoscale.

Using an array of techniques ranging from molecular engineering and thin film technology to the mechanical reduction of bulk materials to ultrafine powders, it is possible to develop materials that have very different physical and chemical properties and combinations of materials that provide unique functionalities of interest.

The Holy Grail would be the definition of a single, “bright-line” description that could distinguish materials and systems that have nanoscale characteristics from those materials and systems that have characteristics that are indistinguishable from existing, well-known, macro-scale applications.

This type of definition could be very simple indeed, with a single, specific dimensional cut-off value identified for “ things nano”. However, the dimensional dependence for the onset of such nanoscale attributes differs from property to property.

For example, the onset of dominant quantum characteristics occurs at dimensional scales of a few nanometres; typically, in the range of up to about 20 nm, nanoscale magnetic characteristics occur up to a maximum of 50 nm to 60 nm while electronic properties show size-dependant variation in a range that extends beyond 1 000 nm.

The variation in dimensional dependence poses difficult challenges in creating terminology that is both accurate (i.e. reflecting physical reality) and precise (i.e. not so vague or general as to encompass such a potentially broad range of materials so as to be relatively useless).

The impact of these differing ranges on clear definitions is significant and requires the development of a carefully structured, taxonomic terminology into which precise definitions can be embedded.

For such a terminology to be useful as a basis for both standards development and as a basis for writing regulatory system entries, it must be searchable in a way that it can provide a basic nomenclature system for nanotechnology.

In turn, such a structured terminology must have the potential for being comprehensive and coherent for nanotechnology terms across the entire field of nanotechnology and must be translatable into existing definitions and terminology across the entire field of standards.

## Differentiating nanomaterials from their macro-usage

It would seem useful to distinguish between nanomaterials and the macro or bulk materials that might contain nanomaterials.

Limiting the definition of nanomaterials to materials that are truly in the nanoscale will achieve clarity in definition and will not in any way limit the ability of industry or governments to evaluate, control or regulate the presence, emissions, discharges, etc. of nanomaterials that may be contained in macro or bulk materials.

Most regulatory regimes recognize this distinction (e.g. not every material that contains a detectable concentration of cadmium is called cadmium).

Similarly, excluding bulk or macro materials from the definition, nanomaterials will not prevent industry from controlling or governments from regulating the presence of nanomaterials in bulk or macro materials.

Bulk materials that, when they degrade, release or discharge nanomaterials can be clearly identified, regulated and controlled, as necessary, without being defined themselves as nanomaterials (e.g. regulations limiting the presence of certain hazardous materials in electronics distinguish between the hazardous substances of interest and the equipment within which they may be found).

It is the distinction between the material of interest and other materials that they may contain that allows industry and government to define more precisely what they want to control and regulate.

## Managing the terminology output

In most cases, the development of standard consensus-based definitions and terminology is derived from a broad usage in scientific and trade literature and the ISO process will draw on existing scientific usage to the extent possible in developing definitions and terminology for nanotechnology.

However, it is relatively early in the development of the field and, as already mentioned, relatively few nanoproducts are on the market. This will require nanotechnology definitions and terminology to evolve as the domain moves towards its full potential. Achieving this dynamic management regime for the definitions and terminology will require efforts outside the normal practices of ISO.

New definitions must be developed regularly and as required during this evolutionary phase, and these will then need to be embedded in the terminology and linked, as appropriate, to other existing standards.

The terminology for nanotechnologies will require ongoing management as it expands and matures over the next decade or so. While it is early to define the nature of such administration, it will be important to recognize these requirements as the definitions and the terminology structure are developed and to incorporate tactics for managing this growth within the ISO process.

## Quotes

**“An important key to the development of nanotechnologies is a coherent and comprehensive vocabulary and terminology.”**

**“Creating a coherent and consistent terminology is important for commercial and general communications.”**

# NEWS

Österreichisches Normungsinstitut
Austrian Standards Institute
Member of CEN and ISO
www.on-norm.at

## Das A-Z der Abwasserbehandlung

### Die neue ÖNORM EN 1085

Wien (ON prm, 2007-06-05) So komplex wie meist der Inhalt von Abwässern – ob aus Haushalt oder Industrie – ist, ist auch ihre Behandlung. Eine Vielzahl von Begriffen wird dabei verwendet, verschiedene Unternehmen und Stellen arbeiten dabei zusammen. Um die Verständigung über Grenzen hinweg zu ermöglichen, liegt nun ein Wörterbuch der Abwasserbehandlung vor – zu finden in der neuen Ausgabe der ÖNORM EN 1085.

Ziel dieser Europäischen Norm ist die Schaffung einer einheitlichen Terminologie in Deutsch, Englisch und Französisch. Die darin festgelegten Begriffe dienen zugleich als Grundlage für die Erarbeitung entsprechender Produkt- und Leistungsnormen und können in speziellen Normen gegebenenfalls präzisiert werden.

Abwasser selbst wird definiert als „Wasser, bestehend aus jeglicher Kombination von abgeleitetem Wasser aus Haushalten, Industrie- und Gewerbebetrieben, Oberflächenabfluss und unbeabsichtigtem Fremdwasserzufluss“. Schmutzwasser hingegen ist die „Summe aus häuslichem und industriellem Abwasser“.

Erhältlich ist
ÖNORM EN 1085 Abwasserbehandlung – Wörterbuch (mehrsprachige Fassung: de/en/fr)

im Webshop
http://www.on-norm.at/shop

## Europäische Bauprodukte

### Neue Europäische Norm für Bauprodukte

Wien (ON prm, 2007-06-05) Die Umsetzung der Europäischen Bauprodukten-Richtlinie ist ein äußerst komplexer Vorgang, der an alle, die an der Erarbeitung der entsprechenden Europäischen Normen beteiligt sind, hohe Anforderungen stellt.

Als Orientierungshilfe wurde dazu nun die ON-Regel ONR 20000 erstellt. Initiiert wurde dieses Regelwerk vom Strategischen Sektorgremium Bauwesen (ON-W 1011).

In der ONR 20000 sind die Terminologie und die wichtigsten bei der Normungsarbeit zu beachtenden Punkte darlegt. Aufgrund der Komplexität der Bauproduktenrichtlinie gibt diese ON-Regel sowohl den Experten in den zuständigen ON-Komitees als auch den mit der CE-Kennzeichnung von Bauprodukten betroffenen Kreisen ein Hilfsmittel für die praktische Umsetzung Europäischer Normen in das nationale Normenwerk.

Erhältlich ist
**ONR 20000** Europäische Normen für Bauprodukte – Begriffe und Umsetzung im nationalen Normenwerk

im Webshop
http://www.on-norm.at/shop

# Hilfe für Menschen mit Behinderungen

## Die aktualisierte Ausgabe der ÖNORM EN ISO 9999 liegt nun vor

Wien (ON prm, 07-06-05) So unterschiedlich die Bedürfnisse von Menschen mit Behinderungen sind, so unterschiedlich sind auch die Hilfsmittel, die zur Verfügung stehen, um diesen Personen das Leben so weit wie möglich zu erleichtern. Die Normung – national, europäisch und international – hat es sich schon vor langem zur Aufgabe gemacht, diese Hilfsmittel mit Hilfe von Standards zu verbessern und weiterzuentwickeln.

Eine wichtige Grundlagennorm dabei ist ÖNORM EN ISO 9999, die dafür die Klassifikation und Terminologie regelt. Sie liegt nun in einer aktualisierten Ausgabe vor. Hilfsmittel (einschließlich Software) sind darin nach ihrer Funktion klassifiziert. Die Klassifikation basiert auf drei Hierarchiestufen, wobei jeder Kode aus drei zweistelligen Ziffern besteht. Wie in sonstigen Klassifikationen werden für jede Stufe Kodes, Benennungen, Erläuterungen, Einschlüsse, Ausschlüsse und Querverweise angegeben.

Neben der inhaltlichen Beschreibung und der eigentlichen Klassifikation sind eine Konvertierungstabelle mit Änderungen zwischen der Vorversion (2002) und der aktuellen Fassung dieser Internationalen Norm sowie ein alphabetischer Index enthalten, um Benutzerfreundlichkeit und Zugänglichkeit der Klassifikation zu verbessern.

Die ISO 9999 wurde 2003 in die Familie der Internationalen Klassifikationen der Weltgesundheitsorganisation (WHO-FIC) aufgenommen. Die WHO-FIC besteht aus hochqualitativen Klassifikationen für relevante Bereiche des Gesundheitswesens. Durch ihre Aufnahme wird die Anwendung dieser Internationalen Norm weiter gefördert.

Erhältlich ist
**ÖNORM EN ISO 9999** Hilfsmittel für Menschen mit Behinderung – Klassifikation und Terminologie

im Webshop
**http://www.on-norm.at/shop**