Guidelines for the Standardization of Terminology for the Welsh Assembly Government Translation Service and the Welsh Language Board

Delyth Prys
Dewi Bryn Jones
2007

Noddir gan
Lywodraeth Cynulliad Cymru
Sponsored by
Welsh Assembly Government

BWRDD YR IAITH
GYMRAEG • WELSH
LANGUAGE BOARD

Contents

1. Background

The first report for the Welsh Language Board on the subject of terminology standardization was written in 1998 which included criteria which were written in 1993 by Delyth Prys on behalf of the Wales Curriculum Authority, but the fast pace of developments in the manipulation and management of knowledge through electronic technologies and the web has meant that much of the information contained in that report has now been supplanted. The Welsh Language Board therefore commissioned this report from the University of Wales, Bangor, in order to update the guidelines and information available in a Welsh context.

1.1 ISO Standards

The report makes frequent mention of the ISO standards. ISO, or the *International Organization for Standardization,* is an independent international body. ISO derives from the Greek word ‘isos’, meaning ‘equal’, and this is the short form of the organisation's name in any language. ISO's main work is the development of technical standards for various fields on a world-wide level. At present, 157 international standards organisations are members of ISO, each state having one vote. The standards body for the United Kingdom and the one which represents its interests and holds the UK's vote in ISO is the BSI (British Standards Institute). Another standards body which exists at a European level is CEN (*Comité Européen de Normalisation*). The members of CEN are the national standards bodies such as the BSI who represent the countries of Europe, and CEN in its turn is affiliated to ISO as the world-wide standards body.

The relevant international standards in the field of terminology standardization derive from the ISO Committee for Terminology and other language and content resources, ISO/TC 37. A list of all the standards for which TC 37 is responsible may be found in Appendix 1. Over the last few years this Technical Committee has extended its remit from its original focus on terminology to encompass other elements pertaining to language management within electronic and distributed environments, as computers and the world-wide web have become more important for managing and distributing knowledge in contemporary society. There was a notable increase of activitiy in this field in Wales as a direct result of the Welsh Language Act of 1993 which created a greater demand for bilingual documents. The role of the Welsh Language Board and its involvement with the translation profession in Wales has also evolved during the same period, especially following the establishment of the National Assembly for Wales in 1999.

The ISO standards on terminology are intended for experts engaged in standards development in every field. They also aid professional terminologists, the authors of multilingual documents, creators of controlled vocabularies (e.g. thesauri), metadata experts and an

increasing number of specialists who deal with knowledge in complex contexts across various linguistic and subject areas. It may seem that Welsh has only a limited role in some of these environments, but adhering to international standards, as far as possible, not only improves the understanding of specialists in the various fields within Wales of the requirements, but also prepares the Welsh language to take advantage of any technological and sociolinguistic developments in the future.

New standards and technologies are both rapidly multiplying in the field of terminology and international language resources. The present report is unable to do more than give a general overview of the field, and summarize the main points of the relevant ISO standards. Any organisation or technical committee involved in a detailed way with any aspect of the work should obtain full copies of the relevant standards and become acquainted with them.

As yet, there have not been any official Welsh adaptations of any ISO, CEN or BSI standards. If these standards are to be recognized and used in Wales, then official Welsh-language versions of the most important and relevant standards for Wales are needed, starting with the standards relevant to terminology work and the translation industry.

1.2 1.2 Explanation of terms

Below are listed the technical definitions of some of the key terms arising in this report. The definitions below are quoted as far as possible from the explanations given in the relevant ISO standards. Many of these terms are defined in ISO 1087-1: 2000*: Terminology work – vocabulary – Part 1: Theory and application*.

Welsh Term	English Term	Definition
canllawiau	guidelines	Guidance given within standards or other official documents on how to attain the objectives for a specific activity. for example, how to choose appropriate terms.
cysyniad	concept	Unit of knowledge created by a unique combination of characteristics.
egwyddor	principle	Basis or general rule aiding in the determination of action or choice between different terms.
gwyddor termau	terminology 2	Science studying the structure, formation, development, usage and management of terminologies in various subject fields.
maen prawf	criterion	Specific principle used to test and measure the acceptability

		of a term.
prosiect	project	Planned work needed to produce a standard that can be agreed upon.
safon	standard	Document established through consensus and approved by recognised bodies, and which provides rules, guidelines or characteristics for general use.
term	term	Verbal designation of a general concept in a specific subject field.
term a ganiateir	admitted term	A term accepted as a synonym for a preferred term by an authoritative body
term anghymeradwy	deprecated term	Term rated according to the scale of the term acceptability as undesired.
term cymeradwy	preferred term	Term rated according to the scale of the term acceptability rating as the primary term for a given concept.
terminograffeg	terminography	Part of terminology work concerned with the recording and presentation of terminological data.
terminograffydd	terminographer	Person who in his or her professional capacity records and manipulates terminological data.
terminoleg	terminology 1	Set of designations belonging to one special language.
terminolegydd	terminologist	Specialist in the science of manipulation and standardization of terminology.

1.3 Technical terms and the Welsh language

It should be borne in mind that the international standards for terminology work and standardization deal with the technical domains of language for special purposes. They do not attempt to pass judgement on which words and phrases should be used in everyday language outside of these narrow confines. Neither do they dictate how gaps in a general language lexicon should be filled, although some of the guidelines on term formation may aid in this work.

Historically, new terms have been introduced into Welsh, as into other languages, as the linguistic community has become aware of new concepts. In a period of substantial change, new concepts, and therefore new terms, flood into a language. This happened extensively with Welsh in the early modern period when the Bible was translated into Welsh, bringing with it new theological terms (words such as ‘iechydwriaeth’ or ‘iachawdwriaeth’ (salvation) where both forms were coined by William Salesbury). This was also the period when new vegetables arrived from the continent of America to Europe, giving us such items as ‘tomatos’ (tomatoes), ‘tatws’ (potatoes) and ‘tybaco’ (tobacco). A number of these words are borrowings through English from Native American languages.

The end of the eighteenth century and beginning of the nineteenth century was another period of rapid change in Wales, not only as the Industrial Revolution gathered momentum, but also as the administration of the modern state developed. This is the period when such terms as ‘pwyllgor’ (committee), ‘gweinyddiaeth’ (administration) and ‘cofnod’ (record) were coined in Welsh. Very often we see these terms appear for the first time in Welsh in the dictionaries of John Walters and William Owen Pughe, with the new vocabulary being disseminated through the Welsh periodical press and translations from the English (Morgan, 2002).

A large number of new concepts came into Welsh with the introduction of compulsory education from the last quarter of the nineteenth century onward. Even before Welsh became a popular medium for education, pressure from the needs of the Welsh language press, popular lectures, and later on, radio and television, meant that there was a demand for appropriate vocabulary in Welsh. A sub-committee of the Language and Literature panel of the Board of Celtic Studies, under the chairmanship of R. Elwyn Hughes, undertook the work of compiling and standardizing lists of technical terms for Welsh. Their principles of standardization were based on the work of Guyton de Morveau *et al*, *Methode de nomenclature chimique* (Paris, 1787), a work which greatly influenced the development of English scientific terms, and which later became the basis for some of the ISO standards (Prys, 2003).

When sports, in particular rugby football, began to be broadcast in Welsh on radio and television, broadcasters such as Eic Davies, and later his son, Huw Llywelyn Davies, endeavoured to coin terms which

were suitable for discussing these subjects, and terms such as ‘mewnwr’ (inside half), ‘maswr’ (outside half), ‘sgrym osod’ (set scrum), ‘cic gosb’ (penalty kick) etc. became familiar words in the language. In contrast to written work, where documents tend to be produced first in English and then translated into Welsh, television and radio sports commentaries are created with Welsh as the original language; this may be the reason for the strength of their natural-sounding Welsh structures and features.

It is a mistake to think that the introduction of new vocabulary to the language is a ‘problem’ which is specific to Welsh. Most of the languages of the world today live in the shadow of English as a dominant language, with concepts, and therefore new terms, reaching them first through the medium of English. In a colloquial and informal register, speakers often borrow from English, adapting English words to conform to the phonetic and grammatical pattern of Welsh, for example, adding the verbal ending '-io' as in 'save' yields 'safio'. However, Welsh speakers, in common with members of other small language communities, have a prejudice against borrowings from the adjacent dominant language and because of this, these informal borrowings are considered to be slang, and an effort is made to find native words, or to borrow from more acceptable languages, such as Latin (Thomas, 1991).

Many borrowings into the general language become respectable at a later stage, and are adopted as technical terms in special language domains (as has happened to terms such as ‘tomatos’, ‘tatws’ and ‘tybaco’ in the field of botany). Many technical terms which came into Welsh originally in exact technical fields have now become part of the general language (for example ‘rhyngwyneb’ (interface), ‘y we’ (the web) and ‘ailgylchu’ (recycle)). Technical terms will become normalized in Welsh if there is a culture of discussing the particular specialist subject area through the medium of the language. This is how the technical terms of the Bible came into common use, in oral and written form, and this is how Welsh rugby terms have become a normal part of the language, mainly through the spoken media.

On the other hand, users may not be aware of terms in another language, either because they are generally monolingual, or because their work or educational environment is close to being monolingual. It may be argued that at least three generations of monolingual, literate Welsh-speakers in the nineteenth century adopted Welsh coinings because they reached them through the Welsh-medium press when they did not have access to English. Schoolchildren today who receive their education through the medium of Welsh learn the Welsh terms if the technical concepts are introduced to them first of all through the medium of that language.

The *Welsh Academy Dictionary* (Griffiths and Jones, 1995) has filled many gaps in the general lexicon of Welsh, as well as in the special language of technical subjects. Without resources such as these, and

the numerous other dictionaries and terminology lists which have been published during the last ten years, the writing of technical documents in Welsh today would be a laborious task, and certainly the work of translators would be much harder. Note however, that these resources have only an indirect influence in terminology dissemination. The vehicles for dissemination are the documents, the lessons, the radio and television programmes which include the terms; it is only then that they gain independent life in the language.

1.4 Terms in a multilingual world

In Wales, the requirements of bilingual policies from public bodies, and the needs of translators, mean that there is a strong bilingual bias (Welsh and English) to the terminology standardization work. This report presupposes that Welsh is the language in which terms need to be standardized. However, the work of standardizing terminology is more effective when it is carried out simultaneously in all relevant languages of the standardization body:

> The preparation of a terminology standard is most effectively carried out simultaneously in all the official languages of the standardizing body. (ISO 10241: 1992 (E) 5.1.6.1)

In practice, in Wales it is usually requested that terminology lists are drawn up bilingually, in Welsh and English. English terms are not usually standardized and this may lead to confusion and inconsistency in the work. Public bodies should consider adopting work processes whereby Welsh and English terms are standardized simultaeneously. Ideally, translation of legislation, regulations and other key documents should be a two-way process, with each language informing the other, rather than a process whereby a document is created in one language and then translated into the other in a linear manner. In a two-way process, there is an opportunity to revise the wording and terminology in both languages in the light of any conceptual problems identified in the translation process.

It should also be remembered that terminology preparation and standardization is a science which essentially can deal with one language at a time, and many of the remarks in this document therefore have the same relevance for Welsh as a language which has its own conceptual space and life, independent of any bilingual or multilingual considerations. However, the international and multilingual dimension is one to be welcomed, not only because it applies the best practices of the international science of terminology to serve the needs of Welsh, but also because it mainstreams Welsh as a language amongst all the languages of the world.

2. Managing terminology projects

2.1 The relevant international standards

The two relevant international standards for reference when operating and managing terminology standardization projects are:

- ISO 15188:2001: *Project management guidelines for terminology standardization*
- ISO 10241:1992: *International terminology standards – preparation and layout*

These standards contain a number of guidelines which recommend the steps to be taken during the life of the project, the output to be produced at each stage as well as the most suitable team structures for various circumstances.

2.2 Appropriate administrative procedures for managing terminology and terminology projects

According to ISO 15188: 2001 (E), in general, there are four phases in a terminology project:

- preparation
- planning
- action
- review

These phases must be followed in their proper order, each one leading on from the other in succession:

> The order of project phases is irreversible. It is based on the principles of terminology and shall not be changed. (ISO 10241:1992 (E) 5.2)

2.2.1 Phase 1: Preparing for the project

The first step is to prepare for the project by (a) holding a feasibility study, (b) drawing up a framework for it and (c) preparing a specification for the work (see ISO 15188:1992 4.2).

a) Feasibility study

When considering the feasibility of the project there is a need to:

- establish the purpose of the project (who needs what and for whom?)
- identify potential users and their needs
- identify situations where there could be a possible risk because of a misunderstanding and where there is a need to harmonize user groups (see ISO 860:1996).

b) Framework

Establishing appropriate procedures includes the legal, financial and organisational details of the project. These are some issues which should be considered:

- the legal aspects of the terminology to be investigated and standardized should be defined. Who has the necessary information? Who has the right to sell it, in part or in full? Who has right of access to the information and when?
- the financial aspects of the project should be defined. Who contributes financially to it? What are the conditions of the contributions? What is the budget? Is the money given in one lump sum or does the budget depend on an analysis of number of terms, human resources etc?
- who are the bodies/organisations who are taking part in the project and what are the conditions of their contribution?
- are any language policies relevant to the work?
- are any national varieties of language being included, e.g. Canadian French as opposed to European French?

c) Specification

The specification should include the results of the project feasibility study, and should establish a framework identifying what is needed for the successful completion of the project. It should take the form of a document which lists, for example:

- the justification for completing the project
- a detailed description of its context and background
- an outline of the provenance and circumstances of the request
- the aims and objectives of the project
- who are the beneficiaries and how they will contribute towards the work, or what will be the conditions of their involvement with the project.

Criteria for accepting the project should also be developed, and they should be compared to criteria formed for similar circumstances. Examples of appropriate criteria include the following:

- the project fulfils users' specific needs
- it comes within the general remit of the proposed workgroup
- it is necessary at this time
- it aims to produce a new product
- it can easily be extended to create further products
- the deliverables are to be distributed in an effective manner
- the project will be accepted by the user group, and there will be a commitment to its implementation.

2.2.2 Phase 2: Planning

ISO 15188:2001 recommends planning the project through undertaking the following tasks (a) select a project manager to lead the work, (b) convene a work group, (c) establish a work plan, (d) establish work methods, (e) establish work tools for the project and (f) establish an order of meetings for the project.

a) **Selection of project manager**
He or she should be familiar with the subject field under consideration, and with the principles of terminology work:

> Given the importance of the role of project leaders in the management of terminology standardization projects, it is important to select a leader familiar with the subject field under study and also with the principles and methods of terminology work. (ISO 15188:2001 (E) 4.3.2)

b) Convening the work group
The group should include the least possible number of members in order to promote effective communication and cooperation. Between 5 and 8 subject specialists are indicated as the usual number for such a committee (4.3.4). The inclusion of an experienced terminologist as a member of the group is also recommended. The work may be organised in a number of ways, with the terminologist either working as a consultant, or taking a lead role. In some situations, it may be useful to differentiate between the responsibilities of a terminologist and a terminographer. The terminographer has responsibility for recording terms and manipulating them, but is not responsible for their standardization. In such circumstances the terminographer is responsible for recording and presenting the terminological data.

ISO 15188:2001 also recommends that the project leader should ensure that all members of the executive committee are familiar with terminology principles and methods. All members of the group should have received basic training in terminology work:

> An introductory tutorial in practical terminology work should be arranged for all the group members. (ISO 15188:2001 (E) 4.3.4)

Cross-referencing with ISO 10241:1992 (E), the possibility of obtaining the assistance of language specialists is also noted. These specialists should be native speakers of the language in the study, or possess a qualification in that language. That assistance is vital in order to create and check definitions, examples, notes and observations:

> Native speakers are essential for the formulation of definitions, examples, notes and comments. (ISO 10241: 1992 (E) 5.1.6.2 (b))

The above summarizes what should happen in an ideal situation. In practice, these guidelines are not always followed (Wright, 2006).

c) **Establishing a work plan**
The plan should be recorded in a Work Plan document. This should include:

- a description of the project structure
- the workgroup membership
- the responsibilities of each member within the group
- the project's work remit
- the timetable (including target dates for each process)
- work methods to be used
- data manipulation tools to be used
- a description of how and when meetings will be held.

It is recommended that an electronic project management system be used to facilitate the management of these tasks within the project (ISO 15188:2001 (E) 4.3.5).

The project remit should also be established as one of the key tasks. More details on this are given in ISO 10241:1992 5, where the following requirements are listed:

- a needs analysis (e.g. who are the target audience, what is the purpose of the proposed work etc.)
- a definition of the boundaries of the subject field under consideration (and therefore what are the sub-fields, and other adjacent fields)
- an examination of the relevant sources in the subject field under consideration (authoritative documents, general documents, pamphlets and reports, experienced people, dictionaries etc.)
- an evaluation of the different sources in order to assess their dependability (it should specify whether or not they are translations, and translated materials should not be given the same weighting as original materials:

  > Translated material shall be used only in exceptional cases. (ISO 10241:1992 (E) 5.1.4.2 (e)))

- a specification of the number of concepts to be examined (it is usually unreasonable to deal with more than about 200 concepts within one project and it should be sub divided into sub-projects if more are involved:

  > Experience has shown that if the number of concepts exceeds approximately 200, a sub-division of the project into a number of sub-projects becomes necessary. (ISO 10241:1992 (E) 5.1.5)

- a specification of the language(s) of the project.

d) Establish work methods (ISO 15188:2001 (E) 4.3.6)

- The project should work with a restricted number of concepts (see 2.2.2 (c) above)
- The project should use principles and methods of terminology work found in ISO 704:2000 (and described in section 5 of this report)
- In order to minimize effort and cost, all relevant previous work in the field should be considered (ISO 15188:2001 (E) 4.3.6).

e) **Establish work tools**

Agreement is needed on the tools to be used to record all the terminology data generated by the project. A description of the tools and recording method should be included in the Work Plan.

The use of a central database is strongly recommended, to facilitate version control and ensure uniformity within and between sets of standardized terminologies (ISO 15188:2001 (E) 4.3.7). Further details on planning the structure of a bilingual and multilingual database may be found in Part 4 of this report.

f) Establish an order of meetings

Meetings should be convened to discuss matters which cannot be solved through correspondence. These meetings should lay the foundations for ways of achieving agreement on concepts, terms and their definitions. This should include the establishment of a concept system, preparation of detailed files with recommendations on the use of the terms, establishment of task forces, brainstorming sessions and discussion groups to resolve specific problems.

As part of the management process there should be an agenda naming the items to be discussed such as problematic concepts, the time allowed for each item, and who will lead the discussion. Sufficient time should be allowed for those taking part to prepare their recommendations.

E-mail messages, web-based discussion groups or video-conferencing should be used to facilitate meetings and speed up the flow of information between those taking part. A database should be kept and constantly updated, containing details of the group members, previous meetings, correspondence and the status of the project. Every member of the group should receive a list of names, addresses, phone and fax numbers and e-mail addresses of the other members.

2.2.3 Phase 3: Action

The action phase includes collating and recording the terminological data using the methods and tools agreed upon in Phase 2: Planning above (ISO 15188:2001 (E) 4.4)

It includes (a) conformity to ISO/TC37 on terminology standardization, (b) following the specification and work plan (c) continuous evaluation of the work and (d) documenting problems and recording decisions.

a) Conforming to ISO/TC37 on terminology standardization
The creation and standardization of terminology lists is discussed in greater detail in section 5 of this report.

(b) Following the specification and the work plan
Ensuring that the work follows the specifications and work plan is the responsibility of the project manager, in consultation with the work group. If needed, the project manager should adjust the work plan as the project progresses. With the assistance of the work group, he or she will compare work progress and outcomes on a regular basis with the original specification and work plan.

(c) Documentation of problems and recording of decisions
The standards do not detail methods of record keeping for discussions and decisions, but note that the source of any information should be clearly shown. They also note that terms and definitions should be updated directly in meetings as far as possible (ISO 15188:2001 4.4).

The project manager should make sure that there is strong evidence to support any comments and that they should aim at resolving difficulties:

> The project leader should insist that comments be well-supported and provide concrete solutions to specific problems. (ISO 15188:2001 (E) 4.4)

Comments should then be summarized with the original text, the comments and the opinion of the project manager being recorded. Group members should analyse the summary, using it to correct problems.

2.2.4 Phase 4: Review and evaluation

This phase includes activity by all the group members in verifying, evaluating and approving the terminological data (ISO 15188:2001 4.5) The outcome is a collection of standardized terms. This collection is evaluated as a *product*. Evaluating the project as a *project* is a separate task.

a) Reviewing the product

The product's quality is measured according to its conformity with the measures outlined in ISO 704:2000, ISO 860:1996, and ISO 10241:1992. This should ensure that the terminological data collection is accurate and complete.

The standards note that it is the subject specialist who should verify the accuracy and clarity of the terminological records, aided if possible by a terminologist:

> Verification of the accuracy and clarity of all entries of the terminological data collection should be accomplished by subject-field specialists, and preferably with the assistance of a terminologist. (ISO 15188:2001 4.5.2)

It is also stated that the technical accuracy of the finalized product needs to be verified, e.g. as to consistency in spelling and any cross-references. Electronic verification by means of computing tools may be a more efficient way of carrying out these formal checks:

> Verification of the terminological data collection may be carried out by computer so as to check the formal requirements of the entries more efficiently. (ISO 15188:2001 4.5.2)

End users should also be asked whether the product fulfils their requirements. Testing should be carried out on the data collection to ascertain:

- completeness
- dependability
- user-friendliness (ISO 15188:2001 4.5.2).

b) Project review

The project's quality is measured according the quality guidelines measurements in ISO 10006:2003. These deal with aspects such as:

- activity management
- task and timetable management
- financial management
- risk identification
- user feedback.

At the project's end a final report should be produced, including a financial review. The final report should summarise the project objectives, analyse each work phase and document the sources used in each phase. This will make it easier to update the work in future, and to plan other similar projects.

The financial review should include a comparison between the estimate of costs made beforehand and the true expenditure. This is considered to be useful information for future projects (ISO 15188:2001 4.6.3).

3. Collecting terms

This section deals with methods of term collection, such as terminology extraction tools and the possibilities and difficulties which arise with their use. It also discusses the types of sources which are deemed to be dependable or which are unacceptable, including web-based research. Reference is also made to the needs of establishing term contexts. The need for translators to collate terms for terminology management as opposed to terminology standardization is also examined.

The most relevant international standards are:

- ISO 10241:1992 (E): *International terminology standards – Preparation and layout* (especially sections 5.2.1, 5.2.2 and 5.2.3)
- ISO 12616:2002 (E): *Translation-oriented terminography* (which gives additional information specifically aimed at translators).

Both standards together enable members of terminology projects to record, maintain and access any data linked to terminology work in an effective and efficient manner.

3.1 Authoritative sources

ISO 10241:1992 (E) recommends that any appropriate source material should be gathered together in order to analyse the terminology in every language used by the project.

Examples of appropriate sources include:

- authoritative documents, e.g. legal documents, regulations, standards etc.
- documents recognised by subject specialists e.g. textbooks, essays, academic papers and volumes
- contemporary sources that may not yet be recognised by others e.g. web pages, pamphlets
- other databases already in existence
- conceptual systems and other similar taxonomies
- dictionaries, vocabularies and encyclopaedias.

Because of the recommendation to use sources for every language, using translation memory is also a suitable source in bilingual projects, in order to ensure term equivalence across two languages.

3.1.1 Evaluating sources

Before beginning to analyse sources of terminological data, each source should be evaluated (ISO 10241:1999 (E) 5.1.4.2) in order to rate it for dependability and authoritativeness. As well as the issues concerning translations outlined above, the following considerations are important. They are independent of the source format:

- is the source contemporary?
- is the author recognised in the relevant subject field?

- is the terminology in the source unbiased (i.e. showing no bias towards a specific school of thought)?
- is the terminology developed according to international guidelines and standards (if pre-existing vocabularies or dictionaries are used)?

The sources or translations will need to be evaluated further in order to decide whether they are sufficiently acceptable and dependable for the project. ISO 10241:1992 (E) recommends that translated sources are only acceptable under special and rare circumstances:

> It is important to determine whether or not the documents used as reference are translations. If they are, the reliability of the translation shall be assessed. Translated material shall be used only in exceptional cases. (ISO 10241:1992 (E) 5.1.4.2)

There are dangers to using translations as dependable sources on points of grammar and idiom. For example, the syntax of the original language may have left a strong imprint on the translation, and it must therefore be rejected as a guide to examining points of grammar. When examining points of Welsh grammar and syntax, the literary work of contemporary Welsh authors who write originally in Welsh may be used as an appropriate source of guidance. The successful entries for the Prose Medal competition at the National Eisteddfod of Wales may be taken as models of high quality Welsh prose; however, use of dialect, slang or other such language registers for literary ends in these works should be noted.

Although technical terms in translated documents may also be non-standard, the paucity of such terms in original Welsh language documentation means that Welsh documents translated from English originals often have to be searched to obtain Welsh candidate terms. In such cases it is important to record whether the source is an original document or a translation, in order to assist the searcher in its assessment.

When using translated material, care should be taken to ensure a balance of different sources. Over dependence on one source, for example, the translation service of an individual agency, can contaminate the data on the frequency of occurrence of a particular term or form. Although web search engines are a convenient way to find terminological information, here again there is a need to exercise caution, as the great bulk of Welsh documents on the web is produced by a relatively small number of translation units. This differs from the situation with English, where the vast scale and amount of material cancels out any bias in data results. Such a bias may still exist in Welsh language documents if a number of them derive from the same translation unit or agency.

At the end of the evaluation exercise, a bibliography list should be supplied, in order to enable every project member to call on any source for their work.

According to ISO 12616:2002 6, recording information on sources is very important for translators. It emphasises the need for appropriate on-going management of terminological sources of information.

3.2 What should be collected

In order to bring together all the concepts relevant to a subject field, and in order to draft terminology lists in the languages chosen for the project, every source identified as an acceptable source for the project needs to be analysed.

The types of information to be collected are:

- terms
- definitions
- synonyms
- antonyms
- context

All information from a source should be gathered in a single exercise/activity (ISO 10241:1992 5.2.1), and the information should be kept in a consistent format throughout the project (*ibid.* 5.2.2).

3.3 Methods of term collection

A number of specialist software programs are available to identify terms within large text sources, if they are available in electronic format. Examples of commercial software products include:

- MultiTrans
- Trados Term Extract
- SQL Server 2005

It is a comparatively simple matter to extract a list of individual words from an electronic corpus. Extracting terms containing more than one word is more challenging. Terminology extraction tools usually work by identifying clusters of words which occur several times throughout the text. For example, if the software sees that the words ‘magnetomedr’ and ‘protonau’ appear together (‘magnetomedr protonau’ (proton magnetometer)) a number of times, then it identifies them as a possible technical term and will include them in a list of candidates for the terminologist.

A list of raw proposals may be too large to process conveniently. Some software programs use ‘stop lists’ of words that it does not want included. In is simplest form, this may be a list of function words and other commonly-occurring strings such as ‘a’, ‘ac’, ‘mae’, ‘wrth’, etc. in Welsh, but it may be added to e.g. through inputting terms already found in the project’s databases. These programs may also be set to omit trade marks, proper nouns and other categories of words not needed for the project.

All the commercial software works well with English, and can mark, along with the term itself, every occurrence where it is used within the corpus. However, there are problems with the ability of such software to identify terms within Welsh language texts. They are unable to cope with Welsh mutations and verb conjugations and therefore the products do not manage to identify ‘fagnetomedr protonau’, as the same term as ‘magnetomedr protonau and may not therefore extract them to the list of possible multi-word terms. A Welsh lemmatizer may overcome this problem, but none are as yet included in the commercially-available terminology extraction tools.

Some programs such as MultiTrans enable the user to extract terms from a translation memory. Despite the warning in 3.2 above about searching translations as reliable sources of terms, searching a translation memory for terms (and comparing the terms in both languages) is a practical method of extracting terms for Welsh language texts.

Another advantage of using translation memory as a source of candidate terms is that it is possible to keep the term context (i.e. the sentence or paragraph where the term occurs) with the term itself. This is essential in order to understand the concept under consideration, and the use of software programs therefore speeds up the process and cuts project costs.

However, in very small projects, or projects where source documents are not available in electronic form, it may be quicker and more accurate to revert to the old technology, and gather terms by hand from the relevant documents. If this is done, it is once again essential that the entire phrase is recorded in order to see the term in its context, and to enter the source for each term.

3.4 Terminology management needs in the translation industry

Collecting terminology is important not only for standardization exercises. The collation of bilingual terminology lists is an important method of managing the use of terms in bilingual organisations, especially in the translation industry. This is done in order to produce a convenient reference list, which helps a translator searching for an unfamiliar term. It is also a way of ensuring consistency, so that terms are not translated in a different manner the next time they are found in a text. Translators need to store and retrieve a much wider range of data than that stored in a traditional terminology database. Amongst these additional elements are:

- expressions
- context
- standard text segments

The format for storing this information should be defined beforehand. The data categories to be included in each record should also be agreed upon.

Most commercial translation memory systems are able to build a bilingual terminology list as the work progresses. Ready-made lists may also be

incorporated into a number of these systems, and the Welsh National Database of Standardized Terminology commissioned by the Welsh Language Board facilitates this (see http:www.welsh-language-board.org.uk/terms)

Storing the terms used by a public body or translation company in a translation memory or similar system helps ensure consistency in the use of terms. However, ensuring consistency and the standardization process should not be confused. Elements of standardization may occur when aiming for consistency, but there is a fundamental difference between a process which extracts terms from translated documents, and a systematic standardization process, as described in the international standards, where concepts are examined as a coherent system within a specific subject field. Terminology standardization is not the role of translators, and Section 2 of this document details the role of subject specialists and terminologists in that work.

However, in small languages such as Welsh, the translator is often the first person who comes across the need for a standardized term in the language. That need is first highlighted because the translator has difficulty in finding an acceptable translation of the term. It the translator does not find guidance by searching in authoritative sources such as general and specialized dictionaries, and technical documents in the language, the translator should work with subject specialists to find ways of expressing the concept:

> In cases where such sources are not available, the translator should cooperate with subject specialists to find ways of translating terminological information adequately. (ISO 12616:2002 4)

The first translation of a term into a language may be very influential, and obtaining a good translation at the first attempt helps to ensure that inferior terms do not gain currency. Influential bodies that are responsible for a great deal of translation work have a special responsibility to ensure that their terms are accurate and acceptable if they are to be published in public documents. Such bodies should ensure that their translators have easy access to appropriate subject specialists as needed and sufficient resources to record and manage their terminological information.

3.5 Sharing terms with other bodies

Collated term lists drawn from their own translated documents can give private translation companies a commercial advantage, and they may not want to share them with other companies. However, the position is different with the translation units in public bodies, and there is much to be gained by sharing their terminology resources freely with other bodies. A notable example of this is TermCymru, the Welsh Assembly Government's on-line terminology database. This provides convenient and free access to the terms used in the Assembly's official business. This is an invaluable source of terms that are not found anywhere else, with the assurance that

they are used in the real world by a body which operates bilingually at a national level.

In such a situation, a method of grading terms is useful to help others decide whether the terms are appropriate for their use. Such a procedure should show whether the term is a candidate which has not been validated (as is inevitable sometimes when the need to submit a translation is urgent), or whether it has been through an internal validation exercise, or has been included in legislative documents and therefore has a statutory status, as it will be helpful for other users in deciding the reliability of the term. Such a procedure may be refined to offer a five point scale, or any number of points to reflect the internal validation mechanisms of the relevant body. However, where there is no statutory obligation for other translators to use the terms, the good name of a particular collection of terms is more important than the status or the grade designated to individual terms, and a three-point scale is usually sufficient to show the main divisions in status.

The experience of the translation services in large public organisations should enable them to identify gaps in the present terminological provision for Welsh. These bodies may then bring their needs to the attention of those who are responsible for the systematic standardization of terms. Translators' requirements may be used as a criterion to prioritize areas in need of intensive standardization work.

4. Structure of terminology databases

This section reports on the standard bilingual terminology database format, naming the fields and field names to be included. The standard format is based on the need to share or re-use data. Examples of field content are given by selecting relevant examples from standard databases. Also discussed are the difference between the form and content of a terminology database maintained for use in terminology standardization work, and the public interface of such a database and a published paper-based list of terms. Guidance is also given on enabling users to search a number of databases simultaneously, using the same interface.

Most terminology projects need a database as a repository to store all the terminological data and as a central resource for the work of team members.

This is acknowledged in the following ISO international standards:

- ISO 12616:2002: *Translation-oriented Terminography*
- ISO 16642:2003: *Computer applications in terminology – Terminological markup framework*
- ISO 12620:1999: *Computer applications in terminology – Data categories*

These standards specify how a database should be structured for terminology projects, and specify standard names for the fields and metadata.

The standards documents referred to above set separate but complementary standards for the design of the database (ISO 16642:2003), and the names and definitions of the fields and metadata (ISO 12620:1999).

Using ISO 16642:2003 and ISO 12620:1999 together allows for the greatest flexibility in responding to the increasing demands of the project. Without these standards, there is a risk that the database may need to be redesigned, or that the terminological entries may need to be revisited during the lifetime of the project.

4.1 Fields and Metadata

Project members will record all the terminological data by entering specific units of information such as '*term*', *'definition*', *'source'* etc.

These specific units are called 'Data Categories' by ISO and ISO 12620:1999, and they include all the different types of Data Categories recommended, including their standard names and meanings.

ISO 12616:2002 recommends the data categories which should be used in terminology databases for translators. The following data categories are mandatory for each term recorded in the database:

Standard name in ISO 12620:1999	**Item no. in ISO 12620:1999**
Main Entry Term:	A.2.1.1
Input Date:	A.10.2.1.2
Source	A.10.19

The remainder of the data categories used are optional and may vary from project to project.

ISO 12616:2002 divides optional categories into three types relevant to different levels of entities in a database, as follows:

Y gwahanol fathau yw :

- Data Categories for terms and term–related information, e.g.
    - Types of terms: *synonym*, *abbreviated form*, *full form*, *symbol*, *variant*, *phraseological unit*
    - Grammar: *part of speech*, *grammatical number*, *grammatical gender*, *noun class*
    - Usage: *usage note*, *geographical usage*, *proprietary restriction*
    - Status: *term status*
    - Equivalence: *degree of equivalence*, *directionality*, *reliability code*

- Data Categories related to concept description, e.g.
  - Domain and sub-domain: *subject field*
  - Concept-related description: *definition,explanation, context, figure, broader concept*
  - Notes
- Administrative Data Categories, e.g.
  - Dates: *input date*, *modification date*, *approval date*
  - Responsibility: *inputter*, *updater*, *approver*
  - Others: *language symbol*, *entry identifier*, *source*

The standards acknowledge that ISO 12616:2002 and ISO 12620:1999 will not be sufficient for some projects. Therefore there is a need to document and share the special data categories used by each project. The following are examples of data category documentation used in Welsh terminology projects:

X.n **Welsh part of speech**
DESCRIPTION: a category assigned to a Welsh term based on its grammatical and gender properties.
PERMISSIBLE INSTANCES: Examples of Welsh parts of speech commonly documented in Welsh terminology databases may include:

Rhan ymadrodd *(Part of speech)*	Byrfodd *(Abbreviation)*
enw gwrywaidd *(masculine noun)*	eg
enw benywaidd *(feminine noun)*	eb
enw gwrywaidd neu fenywaidd *(masculine or feminine noun)*	eg/b
Berfenw *(verb-noun)*	be
adferf (adverb)	adf
adjective (ansoddair)	ans
enw lluosog (plural noun)	ell

The parts of speech shown above are suggestions only. An individual project may decide that there is a need for other parts of speech, such as count nouns and mass nouns. Additional considerations not documented in the standards are the need for consistency in the use of abbreviations to denote parts of speech. For example, the abbreviation **b** has been used in the past to denote both feminine nouns and verb-nouns in Welsh, leading to ambiguity when combining different databases. The above pattern avoids using such ambiguous abbreviations.

4.2 Database structure

ISO 16642:2003 is a complicated standard aimed at language engineers, developers of terminology management systems and data modellers.

Until the revised ISO 12618: *Computer applications in terminology – Design, implementation and maintenance of terminology management systems* is published, ISO 16642:2003 is the only standard available to help terminology projects to structure databases correctly so that they are compatible with the data categories of ISO 12620:1999.

The main purpose of ISO 16642:2003 is the creation of a standard structure (called the Terminology Markup Framework (TMF)) for the various structures used by terminology projects. It allows projects to design their own structure but to do so on condition that they map back to the meta-structure of TMF ISO 16642:2003.

Using the meta structure as a guide for the structure of specific terminology projects, it will be possible to record and relate entities that have data categories to their translations and meaning.

In keeping with the other ISO standards for terminology, ISO 16642:2003 is concept-based. Therefore every terminology entry in the TMF has a concept as its starting point. Starting from the concept, the terminological record includes the various terms in different languages for that concept.

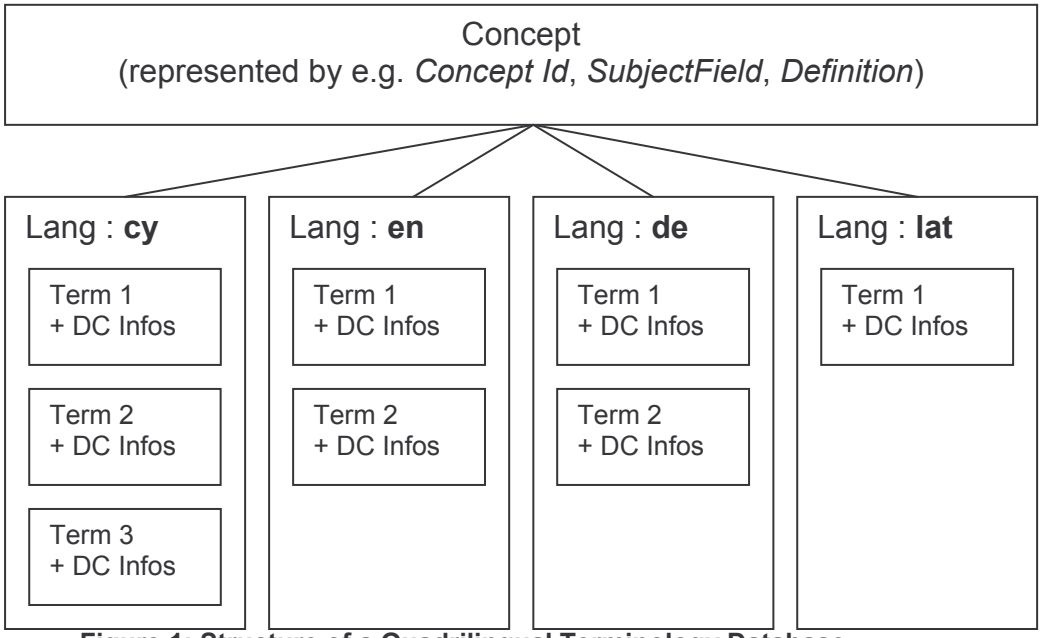


Figure 1: Structure of a Quadrilingual Terminology Database

In this structure, every candidate term in every language is listed, including synonyms. This follows another important principle in ISO 16642:2003, namely that of term autonomy.

An example follows:

Concept:
ID: 6794
SubjectField: Vertebrate animals – Frogs

Language = en (English)

Main Entry Term: edible frog
Input Date: 12/02/2006
Source: *Oxford Nature Terminology Dictionary*
Normative Authorization: preferred term

Language = cy (Welsh)
Main Entry Term: broga bwytadwy
Welsh Part of Speech: eg
Input Date: 16/02/2006
Source: Geiriadur Natur
Normative Authorization: preferred term

Main Entry Term: llyfant bwytadwy
Welsh Part of Speech: eg
Input Date: 16/02/2006
Source: *Y Naturiaethwr*, January 2006, p. 56.
Normative Authorization: admitted term

As seen in the above example, when each term is entered separately in the language, the independent nature of each term makes it possible to record every synonym, abbreviation etc., along with a full collection of data categories.

In particular, data categories that qualify fully the term status as a preferred term, deprecated term or as an admitted term (synonym) can be used – Normative Authorization (A2.9.1). Values that can be assigned to the Normative Authorization data category include:

preferred term (term cymeradwy)	a term recommended by an authoritative body
admitted term (term a dderbynnir)	a term accepted as a synonym for a preferred term by an authoritative body
deprecated term (term anghymeradwy)	a term rejected by an authoritative body

This will also mean that the users of the database can find the 'less standard terms' (such as synonyms and deprecated terms) in the database.

Not every project will wish to record terms in such detail, and may not wish to enable users to search for synonyms and deprecated terms in their databases. It is possible therefore to use data categories listed in ISO 12616:2002 A.2.1 'Types of terms'.

4.2.1 Plurals

Plurals are important data items in Welsh terminology work. The standards do not specify any particular treatment for plural forms of terms, apart from offering methods of modelling and recording parts of speech for every word in the term (this is called the *Term Component List*).

It is possible to create a new data category called 'Welsh Plural' but it should be borne in mind that some nouns may have more than one plural. If the data category 'Welsh Plural' is used, then in order to conform as closely as possible to the standards, different plurals should be separated within the field by a comma, e.g.

Language = cy (Welsh)
Main Entry Term = ci glas
Welsh Part of Speech: eg
Input Date: 16/02/2006
Source: Geiriadur Natur
Welsh Plural: cŵn glas, cŵn gleision

4.2.2 Practical use of ISO 16642:2003

For practical and less abstract use which integrates better with terminology software it is possible to use standards for formats which are available as XML. It is also possible, under some conditions, to use a single database table.

XML

XML standards which conform to the TMF model may be found in ISO 16642:2003. The main standards are TBX and OLIF.

TBX is used within a number of new European and international projects e.g. within EuroTermBank.com (http://www.eurotermbank.com) and also in the National Terminology Database in Wales (Jones and Prys 2005).

An example of data in the form of TBX may be found in Appendix 2.

OLIF (Open Lexicon Interchange Format) (http://www.olif.net/)
The OLIF format standard is also popular. It is a simple format which supports the inclusion of general terminological and dictionary data. OLIF is extensively used in the software translation industry and within machine translation systems.

An example of data in an OLIF format may be found in Appendix 3.

Database Table

A 'flat' terminology database is no longer considered adequate for supporting a structure which conforms to the TMF. But it is still possible to handle the conceptual structure within a simple table structure if the material is simple and completely consistent.

In this design, a row represents a concept – this was described in section 2 of the *Report on the Standardization of Terms Project* prepared in 1998 for the Welsh Language Board. The table has a full collection of terminology data categories for every language mapped to the fields in the database, e.g.

Concept:
ID: 6794
SubjectField:Vertebrate animals – Frogs

Language = en (English)
Main Entry Term: edible frog
Input Date: 12/02/2006
Source: *Oxford Nature Terminology Dictionary*

Language = cy (Welsh)
Main Entry Term: broga bwytadwy
Welsh Part of Speech: eg
Input Date: 16/02/2006
Source: Geiriadur Natur

Main Entry Term: llyffant bwytadwy
Welsh Part of Speech: eg
Input Date: 16/02/2006
Source: *Y Naturiaethwr*, January 2006, p. 56.

When there is more than one term it is possible to divide the field into sub-fields by separating them with a semicolon ';'

Concept Id	Main Entry Term (en)	Input Date (en)	Normative Authorization (en)	Main Entry Term (cy)	Part of Speech (cy)	Normative Authorization (cy)
6794	edible frog	12/02/2006	preferred term	broga bwytadwy; llyffant bwytadwy	eg;eg	preferred term; admitted term

OLEW
The Language Technologies Unit at the University of Wales, Bangor has developed a hybrid format of OLIF, TBX and other formats which is compatible with TMF and with the latest draft version of the Lexical Markup Framework ISO 24613. It also has the flexibility to be used with the ISO 12620:1999 data categories.

OLEW operates as a higher-level structure to assist the dictionary needs of The Language Technologies Unit at the University of Wales, Bangor including terminology, thesauri, general dictionaries, place-names, word lists, pronunciation lexicons and machine translation.

4.3 Production of Electronic and Paper-based Dictionaries

If TMF guidelines on structure are followed, generating different views of the data is a straightforward matter in order to produce both electronic dictionaries and traditional paper-based ones. The back-end is a single standard database, the only change is to the external representation of the interface, e.g. in order to hide or display some fields, or to change the order of the languages shown.

It is also possible to produce paper dictionaries by drawing terms into one table and using the 'create report' features in office software such as Microsoft Office and OpenOffice.org.

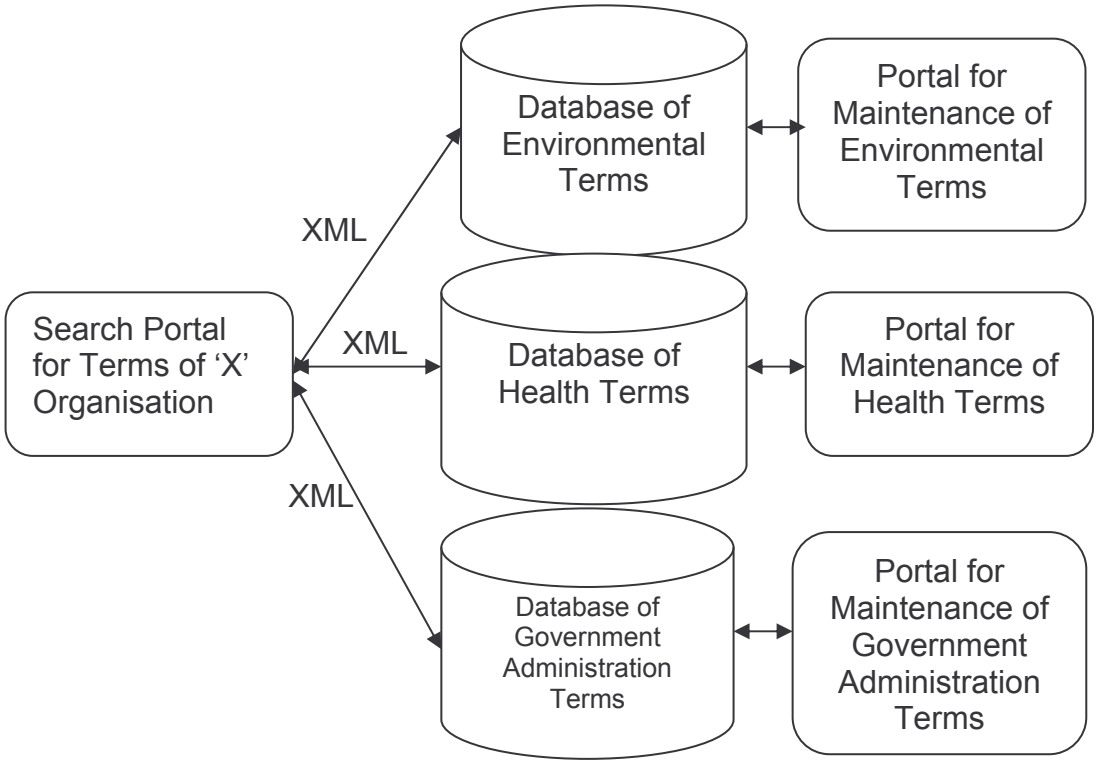
When presenting terms to the ordinary user, some data categories that they do not need to see, such as those categories which deal with project management or internal metadata, may be hidden or omitted.

4.4 Searching in several databases simultaneously

Usually, on-line terminology databases offer an interface for users to enable searches. If the user needs to search in more than one database they must use a specific, separate interface for each database. The search process may therefore become cumbersome and ineffectual.

One of the purposes of the ISO terminology standards is to ensure that terminology databases are interoperable and are able to interact. This facilitates the exchange and inclusion of data between several different databases at the same time, on behalf of the user.

If a database offers an interface for other systems to provide its data in an ISO compatible XML format such as TermBase eXchange (TBX) (http://www.lisa.org/standards/tbx/) or the Open Lexicon Interchange Format (OLIF) then it is possible to include it within a federated network of databases, searchable by websites and other language software.



Without the use of these XML-based standard formats, it is not possible to connect different databases together in order for them to share a search portal. Users cannot create their own search portal, however an organization wishing to make their database interoperable

may do so with its own databases, and/or other databases that are publicly available on the web. There are two conditions to this: firstly, the technological condition that the databases should be in an ISO compatible XML format so that it is technically interoperable, and secondly, the legal requirement that the databases should be free from any copyright or commercial restrictions. The SOAP (Simple Object Access Protocol) conventions (http://www.w3.org/TR/soap/) can be used to implement and control the access rights and interoperability that follow from the editorial, legal and copyright issues of placing such databases on the web.

These TBX and OLIF standards are still relatively new, and as yet there are few examples of interoperable terminology databases on the web. However, the ISO standards now recommend the networking of interoperable terminology databases, rather than the previous approach of collating terminologies into one very large central database. An example of the latter is Eurodicautom: (http://ec.europa.eu/eurodicautom/Controller) which is the European Union's huge database, enabling searches between any pair of languages among the eleven languages which were originally the only official languages the EU (as well as Latin). Following the expansion of the EU to include other countries and languages, and the desire for various EU institutions to share their terminological resources, a new database was developed. The new database is called IATE: Inter Active Terminology for Europe, (http://iate.europa.eu/iatediff/) and the old material from Eurodicautom and other databases was ported to it. However, this database still does not follow the new standards to make databases interoperable, since the project started in 1999, before the new standards were published.

An example of a project which conforms to the new standards and which makes terminology databases interoperable is EuroTermBank. This is a European project aimed mainly at the terminology needs of Estonia, Hungary, Latvia, Lithuania, and Poland as they became part of the European Union (see http://www.eurotermbank.com/). In this case, the website acts as a portal, searching internal and external databases to find translations of terms between a number of different languages, using one comprehensive interface. It is likely that this type of interface, using one search portal to access a number of different databases, will become more common in future as the relevant TBX and OLIF standards have since been published.

5. Criteria for terminology standardization

This section discusses the criteria to be used in terminology standardization, based on the latest developments in ISO 704:2000 and ISO 860:1996 and other international standards, and their application to Welsh.

5.1 ISO standards 704 and 860

ISO 704:2000 is the core international standard dealing with the theory of terminology standardization (Wright, 2006). The standard was applied to Welsh in 1993 (Prys, 1994) and used as a basis for *Y Termiadur Ysgol* (Prys and Jones, 1998) and *Y Termiadur* (Prys and Jones, 2006). This standard was also taken as the basis of criteria for standardizing technical terms in Welsh in the *Report on the Standardization of Terms Project* (Prys and Jones, 1998). These are the criteria which the Welsh Language Board's Standardization of Terminology Team refer to when standardizing terminology. The 1987 version of ISO 704:1987 (E): *Principles and methods of terminology* was used for all these. In 2000 a revised version of the standard was published: ISO 704:2000(E): *Terminology Work – Principles and methods.* This standard was adopted verbatim as a British standard by the BSI, and is therefore operational as a standard for the United Kingdom.

A draft version of a further revision of ISO 704 was produced in 2006, but since that has not yet been approved, it is not quoted here. It is appropriate to note however that revision of the ISO standards is a constant process, and users of standards in Wales should be aware that there may be further revisions of this standard in the future.

ISO 860:1996 is very short, and considers the way in which concepts correspond to one another across different subject areas within the same language, and across different languages. Most of its principles are reiterated in ISO 704:2000.

5.1.1 The changes between ISO 704:1987 and ISO 704:2000

There is no substantive change between the 1987 and 2000 versions. The main changes are an attempt to explain terminology science more fully, detailing and expanding on the examples, and sorting the material into more ordered sub-sections. The practice of quoting English, French and Russian examples within English documents was abandoned, and now only English examples are given. This is indicative of a more general movement away from using these three languages as international languages in the ISO standards. Following the disintegration of the Soviet Union, the contribution of Russian to the standards has disappeared, leaving English and French as the working languages of the committees. It is foreseen that Chinese will increase in importance by the next published version of ISO 704, and this will move the emphasis away from Euro-centric languages and those written with the Roman script or other alphabet-based scripts.

5.1.2 Working conceptually

ISO 704: 2000 once again emphasises that terminology standardization work depends on the concept as the basic unit in the science of terminology:

> Producing a terminology requires understanding the conceptualization that underpins human knowledge in a subject area. (ISO 704:2000 (E) 5.1)

This is the reason why the standards on storage and handling of terminology entries use the *concept* as the unit which anchors the other elements in the entry (see Section 4 for further information on this). This is essentially different to the way in which traditional lexicographers arrange lexicographical records on the basis of the *word* or unit of words. Therefore, 'llygoden' (mouse) would, to lexicographers, be a single record based with meaning 1 (small furry animal) and meaning 2 (computing accessory). To the terminologist, however, they are two quite separate conceptual records: one belonging to the subject field of animals and mammals, and the other to the subject field of computing and hardware. It is essential to understand this difference. The conceptual method of working, as well as the fact that terminology standardization is a prescriptive science (determining which of several candidate terms should be accepted as the standard term) while lexicography is a descriptive science (recording language as it is in all its richness and diversity) is the great difference between terminology and lexicography work. Working conceptually should be part of the terminology training for translators and others who use terminology and/or general language dictionaries, as these differences are not always obvious from the external format of published dictionaries.

Section 3 on Collating Terms discusses the various ways of classifying concepts according to either the detailed relationship of a conceptual cluster with sub-concepts in a hierarchical classification system, or in broader clusters, to create terminology dictionaries for the use of translators and others. Whichever method is chosen to collate and store the terms, what ISO 704:2000 and ISO 860:1996 give us are further guidelines on concepts, their features, their relationship with each other and with different conceptual systems:

> ISO 704:2000 has become the guide of choice for concept modelling and definition writing across the spectrum. (Wright, 2007)

5.1.3 Concept systems

It may help to think of a term as a word label which is given to a particular concept. This word label therefore names or denotes a concept. Concepts are items of knowledge, and in specific subject fields they are related to one another in a structure which forms the knowledge mass of the subject field. The terminology of a subject field is not a random collection of terms, and the concepts should form a rational system based on the relationships between the clusters:

> The *concepts* shall constitute a coherent *concept system* based on the relations established between *concepts*. (ISO 704:2000 (E) 5.5.1)

Many types of conceptual systems (generic, relational, associative) are illustrated in ISO 704:2000, but very little work has been done so far on

developing any native Welsh language concept systems. This may be the result of the substantial cost of this type of work, but another reason is that Welsh, in technical areas, has lived in the shadow of English, and the need for standardized Welsh terms has often arisen in the wake of translation from the English. In such a situation, mapping concepts between the two languages becomes important, and advantage may be taken of concept organization carried out for the other language, or in a multilingual or international situation.

Examples of this work being done in Welsh are, firstly, the work of Cymdeithas Edward Llwyd in developing Welsh names for species in an international taxonomy system. Secondly, *Y Termiadur* followed the systematic classification of chemistry terms as given in *Signs, Symbols & Systematics* (Swinfen, 2000). Thirdly, *Terms for Health Promotion* (Prys, 2000) followed the conceptual classification of the *Multilingual European Thesaurus on Health Promotion* (1996), while *Dictionary of Terms for Woodland Management* (Pommerening and Prys, 2005) followed the international classification of Silvaterm, (http://www.iufro.org/science/special/silvavoc/silvaterm-database/). These efforts work best in areas where similar systems exist in different countries, and where the concepts therefore match one another closely.

The ISO 10241:1992 recommendation that terms should be standardized at the same time in every official language within a territory is also relevant. Therefore, if terms are to be standardized for Welsh, an investigation needs to be carried out to ascertain whether the concepts being considered have already been organized systematically for English, or for a multilingual environment. If they have not, the work can be carried out jointly for the needs of Wales in both Welsh and English, thereby saving money and effort.

The starting point for building a conceptual model is the collation of an unstructured group of concepts belonging to the same conceptual field. This may be used to create a vocabulary in a specific subject field. The steps to be taken are:

- the selection of a concept field, the preliminary denotations and the concepts to be dealt with, taking into consideration the subject field and the user group and their needs
- an analysis of the relationship between the concept and wider and narrower concepts in the hierarchy
- the establishment of the position of these concepts within the concept system
- the formation and evaluation of definitions for the concepts based on their inter-relationship
- the determination of designations for each concept.

5.1.4 Monosemy

Monosemy is the correspondence of one designation to one concept only. In the general language, one word or phrase may denote a number of quite different concepts, or a number of related concepts (e.g., the word ‘mole’ in English denotes a small animal that burrows underground, a type of spy, a unit in chemistry, a breakwater, etc.)

In Welsh, one result of borrowing foreign words is that a word or phrasal unit may enter the lexicon to express a particular concept, and will be accepted for that concept, as that may be the most appropriate designation of the options presented (e.g., ‘apwyntiad’ for the concept of an arrangement to meet someone, i.e. ‘appointment’ in English). But accepting such a borrowing seems to legitimize the word as an item in the language, and the word is then used by extension for other concepts where other designations already exist, for example using 'apwyntiad' also for appointing somebody to a position, where the designation 'penodiad' was already in use. In a technical register, care should be taken that borrowings which denote more than one concept in English or any other language do not expand to denote the same range of concepts in Welsh, where designations for those other concepts already exist. Other examples of this tendency are the use of ‘recordio’ instead of ‘cofnodi’ for the English verb ‘record’ (although ‘recordio’ is acceptable for recording music), using ‘theori’ instead of ‘damcaniaeth’ as a count noun for the English noun ‘theory’ (although ‘theori’ for theory as a mass noun is acceptable), and using ‘esiampl’ instead of ‘enghraifft’ for the English noun ‘example’ (although using ‘esiampl’ for 'example as a pattern of behaviour’, rather than as ‘a specific instance of something’ is acceptable).

5.1.5 Forming definitions

This section gives a brief description of the main principles of writing a definition of a term, as given in ISO 704:2000. The definition is a statement of the concept's meaning, including the subject (the term), the copula (a form of the verb 'to be'), and a predicate (description). The example given in the standard in English is: ‘[A] lead pencil [is a] pencil whose graphite core is fixed in a wooden casing that is removed for usage by sharpening.’ Although the periphrastic form of the verb ‘bod’ ('to be') may be used to express this in Welsh (‘mae pensil plwm yn bensil lle mae'r craidd graffit wedi'i osod mewn casyn pren sy'n cael ei naddu er mwyn ei ddefnyddio’), this may be more succinctly expressed by using the concise form of the verb (‘pensil plwm yw pensil lle mae'r craidd graffit wedi'i osod mewn casyn pren sy'n cael ei naddu er mwyn ei ddefnyddio’). In practical terms, typographic conventions such as a colon, a dash, or beginning a new line in the text may introduce the predicate, instead of using the verb 'bod' as a copula, e.g.

pensil plwm

pensil lle mae'r craidd graffit wedi'i osod mewn casyn pren sy'n cael ei naddu er mwyn ei ddefnyddio

Care should be taken not to write circular definitions, where one concept is defined by using a second concept, and the second concept is defined through using the term or elements of the term that denote the first concepts, as they do not add to our understanding of the concept e.g.

evergreen tree

circular definition	tree with evergreen foliage
corrected definition	tree that retains its foliage throughout its lifetime

Incomplete definitions and negative definitions (what something is not) should also be avoided as far as possible.

Incomplete definitions are those that contain irrelevant features or those that are not crucial to the concept. They can be definitions which are too broad in scope (e.g. describing a mechanical pen as a ‘writing instrument composed of a barrel and a refill’, when that is also true of other writing implements such as felt pens and ballpoint pens), or definitions which are too narrow (e.g. describing a mechanical pencil as a ‘writing instrument composed of a barrel, a lead refill and push-button advance mechanism’, which leaves out mechanical pencils which use other mechanisms for moving the graphite forward). In this case the correct definition would be a: ‘writing instrument composed of a barrel, a lead refill and a lead-advance mechanism’.

Note however that definitions given in laws and regulations often interpret terms rather than defining them in the technical sense given in the International Standards, e.g. in a regulation, a definition of an organization may be given as: ‘for the purposes of this regulation, bodies not operating for profit’. This is not a complete definition of the concept, but an interpretation of it. In a complete definition, for-profit organizations would also usually be included in the concept of an 'organization'. Those who are responsible for forming and interoperating legal definitions should bear in mind that their legal definitions are interpretations, rather than definitions in the sense given to technical definitions in the ISO standards.

Negative definitions are definitions which state what a concept is not, rather than what it is. For example:

deciduous tree

inappropriate negative defnition	tree other than an evergreen tree
corrected definition	tree that loses its foliage seasonally

However, sometimes the absence of a feature is essential to understand the concept, and in such cases a negative definition may be required.

5.1.6 Methods of term creation

The principle of conforming to the norms of individual languages is used when describing how to form new terms. New terms should not be coined unnecessarily, and terms which already exist should be used, if their conceptual equivalence is close enough in the relevant language. If that is not possible, then there is a presumption in favour of native forms over foreign borrowings, although borrowing foreign words is not disallowed.

Appendix A of ISO 704:2000 (E) gives examples of methods of term creation for the English language with a warning that this should not be translated to other languages but that it should instead be adapted to the specific rules of the language under consideration. The appendix indicates three methods of term formation for English, noting that other methods may also be possible. The three methods are a) creation of new terms, b) use of forms which already exist, c) borrowing from another language. These three methods are equally relevant to Welsh, and possible Welsh examples are given below as well as the English examples given in the ISO document.

a) Creating new forms

This means creating new entities that did not previously exist in the lexicon. New forms may be created through the following processes:

- derivation
- compounding
- abbreviated forms

i) Derivation

In the derivation process, new terms are formed by adding one or more morphological elements, for example a prefix or a suffix, to the word root.

English example:
phosphor+ous = phosphorous

Welsh example:
blaenoriaeth + u = blaenoriaethu

ii) Compounding

Compounding entails combining already-existing words or word elements to create a new form, which includes two elements or more, but which denotes just one concept.

English example: outflow

Welsh example: gweithgynhyrchu

In Welsh, both strict compounds (e.g. 'byrdymor' (short term) and loose compounds (e.g. 'tymor byr') are allowed. The international standards do not give guidance on this question, but as further guidelines for Welsh, where the same word elements exist in both strict and loose compounds, both forms should be accepted as variants of the same term. In such cases, it is the language register which distinguishes between them. Strict compounds tend to be used in more formal terms, in a high register, with loose compounds being closer to a general, every day language register, and therefore more acceptable in non-technical registers. Note that this sometimes conflicts with the desire for concise, short forms and a word-for-word correspondence, if the corresponding term in English only contains one word.

iii) Shortened forms
Shortened forms may include short forms, abbreviations, use of initial letters and acronyms. These methods are not as popular in Welsh as in English, partly because mutations at the beginning of words may disguise a short form, and partly because common letters in Welsh (such as the letter 'c') arise often in acronyms, so that it is not possible to distinguish between them.

English example of an acronym: UNESCO (United Nations Educational, Scientific and Cultural Organization)

Welsh example of an acronym: CBAC (Cyd-bwyllgor Addysg Cymru)

b) Use of pre-existing forms
Forms which already exist may be used to create new terms through processes such as semantic transfer and cross-disciplinary borrowings. Care should be taken lest the use of pre-existing forms leads to homonymy (the same designation being used for different concepts), thereby leading to confusion and lack of clarity.

An English example of a cross-disciplinary borrowing:
virus (medicine) = infectious agent which causes diseases;
virus (computer science) = infectious agent causing computer malfunction

A Welsh example of a cross-disciplinary borrowing:
ystod (agriculture) (in English ‘swath’: measure of grass that may be cut with one swing of the scythe);
ystod (education) (in English ‘range’: measure of the extent of ability or age of children in one cluster)

In the English version of ISO 704:2000, an example of a change in the semantic category of a word is also given, e.g. for the adjective 'empty' in the field of recycling, the verb 'to empty' is given, and also the noun 'an empty'. In Welsh, other semantic categories use different suffixes,

and an example of this was given in a) i) above, where a noun is transformed into a verb (blaenoriaeth+u).

c) Borrowing from another language
This may happen through a direct borrowing or through literal translation.

An English example of a literal translation:
(German) Raster → (English) raster (digitizer grid)

A Welsh example of literal translation:
(English) trefoil headed pin → (Welsh) pin pen meillionen

In many languages there is a prejudice against borrowing from a dominant language (which is perceived to be a threat to the indigenous language), while it is seen as more acceptable to borrow from other languages perceived to be either neutral or of higher status (Thomas, 1991). This is true of Welsh, where there is often prejudice against borrowing from English, but where learned borrowings from Greek or Latin are acceptable. Compare this situation with the Breton language, where there is a prejudice against borrowing from French, and where Latin borrowings are also regarded with suspicion, as it is often not easy to distinguish between them and words of more recent French provenance. The aim of the international standards is to promote effective communication, rather than to preserve linguistic purity, but giving due regard to the wishes of the specific linguistic community is part of any comprehensive terminological planning (Galinski, 2005).

See also considerations of monosemy in 5.1.4 above.

5.2 Applying the Criteria for Welsh

In the *Report on the Standardization of Terms Project* (Prys and Jones, 1998) for the Welsh Language Board, the ISO 704:1987 (E) criteria for selecting appropriate terms were quoted as follows:

- a term should be linguistically correct
- a term should reflect, in as far as possible, the features of the concept given in the definition
- a term should be concise
- a term should be able to generate other forms
- a term should correspond to one concept only

These remain valid in the 2000 version of the standard, with new additions and details. Below is a summary of the principles of term formation as given in the 2000 version.

5.2.1 Principles of term formation

The patterns of term formation depend on the lexical, morphosyntactic and phonological structure of individual languages, and therefore these should be discussed in national or regional standards dealing with a

specific language, rather than in International Standards. A concise guide for Welsh is given in Section 5.1.5 above.

Pre-existing terms which are well-established in the language should be respected, even if they do not conform fully to the forms of the language, unless there are very sound arguments for change:

> Established and widely used *designations*, even if they are poorly formed or poorly motivated, should not be changed unless there are compelling reasons. (ISO 704:2000 (E) 7.3.1)

When there is more than one designation for a single concept, the one fulfilling the greatest number of criteria in this section is the one that should be chosen:

> If several *designations* exist for a single *concept*, the one that satisfies the largest number of principles listed below should be selected. (ISO 704:2000 (E) 7.3.1)

Therefore there is no hierarchy of principles to aid the resolution of dispute when different criteria are in conflict, and no one criterion is more important than the others.

5.2.2 Transparency

Where possible, something of the meaning of the concept should be inferred from the term that represents it, without the need for a definition. In the ISO standard an example is given of choosing between two designations: ‘torque wrench’ or ‘monkey wrench’ to denote a type of wrench which measures turning power. It states that ‘torque wrench’ is the best term because the ‘torque’ element explains something of the objects' purpose (turning power), while ‘monkey’ (derived from the name of the inventor Möncke), is obscure in meaning.

5.2.3 Consistency

The terminology of a specific subject field should not be an arbitrary and disjointed collection. The terminological system should be rational and cohesive, corresponding to the conceptual system. New terms should integrate into the pre-existing system. For example, synthetic materials such as nylon, orlon, dacron, rayon, etc. all end with the suffix ‘–on’. Therefore if a new synthetic material is being invented, its name should conform to the same system and naming conventions.

5.2.4 Appropriateness

Candidate terms should conform to the usual pattern of meanings within the language community. Confusing terms should be avoided. For example, the English term ‘atomic energy’ should be avoided because it is misleading, suggesting as it does that the energy is formed from the atom. The term ‘nuclear energy’ is more scientifically correct and precise.

Terms should also be as neutral as possible. They should avoid any unfortunate tendencies or negative connotations. The example given is that of replacing the English term ‘genetic manipulation’ with the less negative term ‘genetic engineering’.

5.2.5 Linguistic economy

A term should be as concise as possible. Long, complex terms, especially those containing five or six words, are unwieldy. However, this principle sometimes conflicts with the need for accuracy in scientifically precise terms.

5.2.6 Derivability

This is the principle which was called ‘the ability to generate other forms’ in ISO 704:1987. This principle is reiterated, stating that productive forms, which allow other terms to be created from them, should be favoured, according to the conventions of the individual language. The English example of deciding between the two competing terms ‘herb’ and ‘medicinal plant’ is given, with ‘herb’ being chosen because the terms ‘herbaceous’, ‘herbal’, ‘herbalist’ and ‘herby’ are derived from it, whereas no other terms derive from ‘medicinal plant’.

Although 5.2.1 above notes that there is no priority given to one criterion over another, and that it is the total score for each candidate term which counts, in Welsh the principle of enabling the generation of other forms has proven to be very important, and has served to clinch the argument as to the most appropriate term on a number of occasions. Failure to choose a generative term may pose problems later on, when the number of related terms in a subject field begin to multiply, quite apart from the need to consider plural forms and other parts of speech in Welsh.

Appendix 4 shows a reproduction of an input form used in an exercise to standardize Welsh terms, where the ability of different candidate terms to conform to these principles may be recorded. Note especially that there is space on the form to record derived forms (in section 4) which may be generated in order to measure the productivity of each candidate term, and to record what those derived forms might be.

5.2.7 Linguistic correctness

'A term should conform to the morphological, morphosyntactical and phonological norms of the language in question.’ (ISO 704:2000).

As this is expressed differently in different languages, the international standards do not elaborate on this principle. We therefore reiterate what was stated in the *Report on the Standardization of Terms Project* (Prys and Jones, 1998) to the Welsh Language Board, namely that *Geiriadur Prifysgol Cymru* (The University of Wales Dictionary of the Welsh Language) and *Orgraff yr Iaith Gymraeg*, 1987edition (the standard reference book of Welsh orthography) should be followed for matters concerning the writing of Welsh.

Deciding on matters of orthography should not be left to terminologists, and a panel of language specialists is needed to undertake this task. The use of hyphens and the use of accents are two problems that occur often when creating new terms. With new vocabulary constantly flowing into the language it would be benefitial to establish a procedure whereby any orthographical issues could be resolved.

We also confirm, as in the 1998 Report, that Peter Wynn Thomas' *Gramadeg y Gymraeg* (University of Wales Press 1996) should be followed as the most comprehensive and convenient reference book for contemporary Welsh grammar.

5.2.8 Principle of precedence for the native language

Although borrowing from other languages is acceptable for term creation, ISO 704:2000 states that priority should be given to terms from the native language rather than direct borrowings:

> […] native language expressions should be given preference over direct loans. (ISO 704:2000 (E) 7.3.8)

5.3 Preferred, admitted and deprecated designation of terms

When considering what designation to give an individual concept in a concept system, very often there will be more than one candidate to choose from. The one that is finally chosen will become the standard term, (called the *preferred term* in the English versions of the ISO standards). Of the other candidate terms, they may be classified either as synonyms that are acceptable (called *admitted terms* in the English versions), or as terms which had been considered but had been rejected as unacceptable (called *deprecated terms* in the English versions).

Terms may be rejected for a number of reasons. These may include:

- the term is also used to denote another concept (the monosemy principle of one term, one meaning)
  e.g. in English the term 'load' has been deprecated as a synonym for the term 'force'. On the other hand, the term 'load' is the preferred term to express the associated concept 'application of a force'
- it is inaccurate or incorrect for the designation of the concept under consideration, e.g. in English the term 'fireproof' is misleading, the terms 'fire resistant' or 'fire retardant' are more correct.

It is also recommended that the reasons for deprecating a term be explained if possible. In the *Report on the Standardization of Terms* Project for the Welsh Language Board, it was recommended that the deprecated terms be recorded in field 'Not this term' (Field 11) in the ideal database. It was also noted that the reasons for rejecting a term could be included in a comments field (Field 13) in the ideal database, but that this field would not necessarily be published. In view of the ISO emphasis on explaining the reason for rejecting a term, we now recommend that a specific field be used to record the reason for rejecting the deprecated term, and that the reason be published where appropriate.

In Welsh, the most serious errors in translation are made where there are homonyms in the source language (words which have more than one meaning) and the wrong meaning is chosen, leading to the wrong concept being represented in the target language. Three examples are given below:

English Term	Mistranslation	Correct translation
concept of the mole	cysyniad y wahadden	cysyniad y môl
language register	cofrestr iaith	cywair iaith
electoral address	cyfeiriad etholiadol	anerchiad etholiadol

Because of this type of serious error, it is recommended that the inappropriate designation for the concept under consideration be included as a deprecated term in a terminological record, in order to warn users against misunderstanding the concept and hence choosing the wrong designation. This is an extension of the example given when discussing 'load' above, but now the term does not refer to a related concept in the same subject field, but to a completely different concept in a different subject field.

5.4 Use of disambiguators

Where a full definition is included in a terminological entry, or where a list of terms is confined to a narrow subject field, the need for disambiguators does not arise. However, in a large terminology list, which usually includes more than one subject field or sub-field, homonyms may be disambiguated by the use of concise descriptions to distinguish between them. This is especially important in bilingual terminology dictionaries used by translators, as they are a convenient and fast way of warning the translator to stop and think about the meaning of the concept, and to choose the correct term.

Disambiguators may be shown in parentheses after the term in the source language. A disambiguator may be a synonym, or a concise definition (shown with the equals sign =), or may be guidance for the type of category it belongs to, with the designation ‘of’, ‘in’, ‘with’ etc. Some examples follow:

English Term	Disambiguator	Welsh Term
grain	(=particle)	gronyn
grain	(=food crop)	grawn
grain	(in rock, wood, cloth)	graen

Where one concept is found in a number of fields, but another concept belongs to a very specialized field, the disambiguator ‘in general’ may be used with the first, and the specialist subject field given as a disambiguator in the second. The English part of speech may also help to disambiguate, e.g.

English Term	Disambiguator	Part of Speech	Welsh Term
cool	(=cold)	adj	oer
cool	(of iron setting, weather)	adj	claear
cool	(in general)	v	oeri
cool	(of weather)	v	claearu

6. Language Registers

Terminology handling and standardization are aimed at the need for effective communication and knowledge exchange in the technical register of different subject fields. It has become increasingly obvious, however, that it is not possible to divorce this register from non-technical everyday language. Ordinary users in many languages complain that official forms and documents are full of jargon or obscure terms. This led to the *Plain English Campaign* in English, and *Cymraeg Clir* in Welsh. These efforts, and similar efforts in other languages, recommend using short sentences, avoiding the use of the impersonal and subjunctive modes, and a number of similar guidelines concerning style and syntax.

It is recommended that technical terms are avoided in these non-technical registers as far as possible, and that concepts are explained in a simple manner to the target audience. However, this does not mean that there is no place for technical terms in a language. Subject specialists have argued that the precision and accuracy that standardized technical terminology gives them in their specialist subject is essential for accurate communication. In reality there is no conflict here; there is room for both technical and non-technical registers in a language. Usually a translator will retain the language register of the source text when translating, but there are occasions where the source language register is unsuitable for the target audience. An example of this situation may be a series of pamphlets written by doctors for patients returning home from hospital. The original pamphlets may have been written using technical language, as the authors were unfamiliar with communicating in a simpler register. In such a case, the translator, with the permission of the person commissioning the translation, may improve the communication of the target language by explaining the technical terms or using less technical synonyms for them. Public signs, television advertisements, and press advertisements are other areas where it may not usually be suitable to use complicated technical terms, but rather a simple language register which is accessible to everyone.

If there is a need to change the language register in a translation project, that should usually be discussed with the commissioner of the work, and it should be specifically incorporated into the project (i.e. record the requirement in the specifications, or in the project's final report).

In the *Report on the Standardization of Terms Project* (1998) there was some experimentation with suggesting less technical terms for use when the standardized terms were likely to be in a high register, with the warning that they were not to take the place of technical terms when the latter would be more appropriate:

> These cannot supplant the precise legal term in a technical context, but can assist in the further step of interpreting terms to the public. (*Report on the Standardization of Terms Project* (1998) 2.7.4)

Because of this step, some terminology dictionaries experimented with suggesting non-technical terms, showing them with F for *familiar term*, side-by-side with the technical standardized terms. It should be noted, however, that these non-technical terms are only pointers, and that these suggestions may be expanded with synonyms, dialectical forms, colloquial forms and so forth according to need.

The international standards provide that such terms may be recorded in the other terminology fields permitted in databases of standardized terms. The responsibility lies with the user to choose an alternative term where the standardized term is not user-friendly. Dialectal or local use of terms may also be noted in such fields, but are not standardized as are the technical terms. The other terms are recorded, to be used according to the needs of different language registers. In documents aimed at the whole of Wales, it may be important, for example, to avoid words which have a dialectal flavour, but in other materials such as local newspapers, dialect terms may add to the atmosphere of the publication.

In reality, this is part of comprehensive general-language dictionary provision, and demonstrates how standards for tagging technical terminology databases and traditional dictionaries are beginning to merge. It is also acknowledged that there is a large degree of interaction between the vocabulary of the specialist language and everyday language:

> The interaction between domain languages and general language is very dynamic. There is a constant flux of lexical material in both directions. (Galinski, 2005)

The current ideal is to have a database template which can handle the whole lexicon of the language, including recording dialect forms and local variations, slang words and any other type of vocabulary desired. It is not suggested that every field available in the database should be populated in every project. However, denoting a general pattern that individual projects can map to is a great step forward towards an interoperable procedure where data may be compared and exchanged between different projects and dictionaries.

Appendix 1: Current ISO\TC 37 standards

Terminology and other language and content resources

TC 37/SC 1	
ISO 704:2000	Terminology work – Principles and methods
ISO 860:1996	Terminology work – Harmonization of concepts and terms
ISO 1087-1:2000	Terminology work – vocabulary – Part 1: Theory and application
TC 37/SC 2	
ISO 639-1:2002	Codes for the representation of names of languages – Part 1: Alpha-2 code
ISO 639-2:1998	Codes for the representation of names of languages – Part 2: Alpha-3 code
ISO 639-3:2007	Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages
ISO 1951:2007	Presentation/representation of entries in dictionaries – Requirements, recommendations and information
ISO 10241:1992	International terminology standards – Preparation and layout
ISO 12199:2000	Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet
ISO 12615:2004	Bibliographic references and source identifiers for terminology work
ISO 12616:2002	Translation-oriented terminography
ISO 15188:2001	Project management guidelines for terminology standardization
TC 37/SC 3	
ISO 1087-2:2000	Terminology work – vocabulary – Part 2: Computer applications
ISO 12200:1999	Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange
ISO 12620:1999	Computer applications in terminology – Data categories
ISO 16642:2003	Computer applications in terminology – Terminological markup framework
TC 37/SC 4	
ISO 24610-1:2006	Language resource management – Feature structures – Part 1: Feature structure representation

Appendix 2: Example of TBX terminology data

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE martif PUBLIC "ISO 12200:1999A//DTD MARTIF core (DXFcdV04)//EN"
"TBXcdv04.dtd">

<martif type="TBX" xml:lang="en">
    <martifHeader>
        <fileDesc>
            <titleStmt>
                <title>IBM Data</title>
            </titleStmt>
            <sourceDesc>
                <p>from an IBM Terminology Sample</p>
            </sourceDesc>
        </fileDesc>
        <encodingDesc>
            <p type="DCSName">SYSTEM "TBXDCSv05b.xml"</p>
        </encodingDesc>
    </martifHeader>
    <text>
        <body>
            <termEntry id="c1">
                <descrip type="subjectField">Software</descrip>
                <descrip type="relatedConcept">navigation tree</descrip>
                <descrip type="relatedConceptBroader">navigation</descrip>
                <langSet xml:lang="en">
                    <admin type="productSubset">Servers: eServer
xSeries</admin>
                    <adminGrp>
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>navigation trail</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                        <descrip type="context">The trail at the top of the
content frame is your navigation trail, sometimes called a branch.</descrip>
                    </ntig>
                </langSet>
                <langSet xml:lang="en">
                    <admin type="productSubset">Servers: eServer
xSeries</admin>
                    <adminGrp>
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>branch</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                        <descrip type="context">The trail at the top of the
content frame in your navigation trail, sometimes called a branch.</descrip>
                    </ntig>
                </langSet>
                <langSet xml:lang="fr">
                    <admin type="productSubset">Servers: eServer
xSeries</admin>
                    <ntig>
                        <termGrp>
                            <term>historique de navigation</term>
                            <termNote type="partOfSpeech">nhom</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
```

```
            </termEntry>
            <termEntry id="c2">
                <descrip type="subjectField">Hardware \ Other Processing
Units and Specialized Devices</descrip>
                <descrip type="relatedConceptBroader">acceptor</descrip>
                <langSet xml:lang="en">
                    <admin type="productSubset">Retail Store
Solutions</admin>
                    <adminGrp>
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>bill acceptor</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                        <descrip type="context">Accepts bill denominations
of $1, $2, $5, $10, $20, $50 and $100. The bill acceptor cassette holds 600
bills. It detects and rejects counterfeit bills.</descrip>
                    </ntig>
                </langSet>
                <langSet xml:lang="fr">
                    <admin type="productSubset">Retail Store
Solutions</admin>
                    <ntig>
                        <termGrp>
                            <term>accepteur de billets</term>
                            <termNote type="partOfSpeech">nhom</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
            </termEntry>
            <termEntry id="c3">
                <descrip type="subjectField">Software \ Application
Infrastructure Services \ System Management</descrip>
                <langSet xml:lang="en">
                    <admin type="productSubset">Software: Tivoli</admin>
                    <adminGrp>
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>scheduled operation</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                        <descrip type="context">Using the CAD to manage the
scheduler requires very little memory between scheduled
operations.</descrip>
                    </ntig>
                </langSet>
                <langSet xml:lang="fr">
                    <ntig>
                        <termGrp>
                            <term>opÃ©ration planifiÃ©e</term>
                            <termNote type="partOfSpeech">nhom</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
            </termEntry>
            <termEntry id="c4">
                <descrip type="subjectField">Software</descrip>
                <descrip type="relatedConceptBroader">gallery</descrip>
                <langSet xml:lang="en">
                    <admin type="productSubset">Other -
Miscellaneous</admin>
                    <adminGrp>
```

```
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>samples gallery</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                        <descrip type="context">The program may contain
sample source code or programs, which illustrate programming techniques.
These samples are found in the samples gallery.</descrip>
                    </ntig>
                </langSet>
                <langSet xml:lang="fr">
                    <adminGrp>
                        <admin type="sourceIdentifier">IBM
publication</admin>
                    </adminGrp>
                    <ntig>
                        <termGrp>
                            <term>galerie d'exemples</term>
                            <termNote type="partOfSpeech">nhom</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
            </termEntry>
            <termEntry id="c5">
                <descrip type="subjectField">General</descrip>
                <langSet xml:lang="en">
                    <admin type="productSubset">Retail Store
Solutions</admin>
                    <adminGrp>
                        <admin type="sourceIdentifier">Translation Services
Center</admin>
                    </adminGrp>
                    <descripGrp>
                        <descrip type="definition">The area of a store
comprising conventional lanes and IBM lanes with a paystation.</descrip>
                    </descripGrp>
                    <ntig>
                        <termGrp>
                            <term>front end</term>
                            <termNote type="partOfSpeech">noun</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
                <langSet xml:lang="fr">
                    <admin type="productSubset">Retail Store
Solutions</admin>
                    <ntig>
                        <termGrp>
                            <term>avant-guichet</term>
                            <termNote type="partOfSpeech">nhom</termNote>
                        </termGrp>
                    </ntig>
                </langSet>
            </termEntry>
        </body>
    </text>
</martif>
```

Appendix 3: Example of data in an OLIF format

The English and German concepts are linked through the same concept number ('ConceptUserId') :

```
<entry ConceptUserId="0731F16CCCD2D3119B4D">
   <mono>
      <keyDC>
         <canForm>table</canForm>
         <language>en</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry>

<entry ConceptUserId="0731F16CCCD2D3119B4D">
   <mono>
      <keyDC>
         <canForm>Tabelle</canForm>
         <language>de</language>
         <ptOfSpeech>noun</ptOfSpeech>
         <subjField>general</subjField>
         <semReading>86</semReading>
      </keyDC>
      <monoDC>
            .....
      </monoDC>
   </mono>
</entry
```

Appendix 4: Example of a term standardization form

Concept ID		Field Tag		English Term	
English Definition					

	Welsh Candidate Term 1	Welsh Candidate Term 2
1. Linguistic correctness		
2. Concept features		
3. Cryno		
4. Generation of other forms		
noun		
Noun plural		
verb		
adjective		
opposite		
other		
5. Monosemy		

Note

Bibliography

C. Galinski, *Guidelines for Terminology Polices* (UNESCO; Paris, 2005).

B. Griffiths a D. G. Jones, *Geiriadur yr Academi/The Welsh Academy English-Welsh Dictionary* (University of Wales Press; Cardiff, 1995).

D. B. Jones a D. Prys 'The Welsh National Online Terminology Database', *Proceedings of the Lesser Used Languages and Computer Linguistics Conference* (Bolzano, October 2005).

M. E. Morgan, *Hanes geiriaduraeth yng Nghymru o 1547 hyd 1914: gyda sylw arbennig i ddylanwad John Walters a William Owen Pughe ar eiriadurwyr 1805-1850* (Unpublished PhD thesis, University of Wales Bangor, 2002).

A. Pommerening a D. Prys, *Geiriadur Termau Rheoli Coetiroedd/Dictionary of Terms for Woodland Management* (University of Wales Bangor, 2005).

D. Prys, *Adroddiad Project Cysoni Termau Cymraeg* (The Curriculum Authority, 1994).

D. Prys, 'Providing the terms: standardizing terms for education in Wales', *Proceedings of SCUTREA 2003*, tt. 192-6.

D. Prys a J. P. M. Jones, *Y Termiadur Ysgol* (ACCAC, 1998).

D. Prys a J. P. M. Jones, *Adroddiad ar y Prosiect Safoni Termau* (Welsh Language Board, 1998).

D. Prys, *Termau Hybu Iechyd / Terms for Health Promotion* (University of Wales Bangor and the North Wales Health Authority, 2000).

D. Prys a J. P. M. Jones, *Y Termiadur* (ACCAC, 2006)

K. D. Schmitz, 'Data Modelling: From Terminology to other Multilingual Structured Content', *Proceedings of International Conference on Terminology, Standardization and Technology Transfer* (Beijing, 2006).

T. C. Swinfen, *Signs, Symbols and Systematics* (Association for Science Education, 2000).

G. Thomas, *Linguistic Purism* (Longman, 1991).

S. E. Wright, 'Standards for the Language Industry, *Terminology, Computing and Translation* (Gunter Narr Verlag; Tübingen, 2006), tt. 19-39.

S. E. Wright, 'The Once and Future ISO 704', *Dokumente edition 2007* (DTT, 2007).

Index