

Guidelines for the development of Controlled Vocabularies

Status of This Document

This document describes the Proteomics Standards Initiative (PSI) and the Metabolomics Standards Initiative (MSI) community practices regarding common features of the controlled vocabularies (CVs) delivered by each working group. Distribution is unlimited.

Copyright Notice

Copyright © 2007 [Proteomics Standards Initiative / Metabolomics Standards Initiative] All Rights Reserved.

Abstract

The Human Proteome Organisation (HUPO) Proteomics Standards Initiative (PSI) and the Metabolomics Standards Initiative (MSI) define community standards for data representation in proteomics and metabolomics, respectively, with the primary aim of facilitating data exchange, comparison and verification.

CVs provide terms and their intended meaning (semantics) to be used by consensus annotation systems to standardize the meaning, syntax and formalism of terms used across proteomics, and or metabolomics terminologies. This document describes the design principles which must be followed in building controlled vocabularies (CVs) under the umbrella of the PSI and MSI. These guidelines are intended to facilitate cross domain CV-referencing, -comparison and -integration.

Contents

Abstract	1
1. Introduction	2
2. Scope of the document	2
3. Notational Conventions	3
4. CV metadata	3
4.1 RA attributes (metadata for the whole CV)	3
4.2 RU attributes (metadata for each CV term)	4
5. Identifier	4
5.1 CV Namespace	4
5.2 RU identifier	4
6. RU Label Naming Conventions	5
6.1. Character set and word separators	5
6.2. Spelling and linguistic convention	5
7. RU definitions	6
7.1 General rules for creating sound normalized definitions	6
8. Source of RU Definitions	7
9. RU Synonyms	8
10. RU Comments	8

11. Relations between RU's	8
12. Obsolescing RU's.....	8
12.1 Alternative terms for obsolete RU's.....	9
12.2. Restoring obsolete terms	9
13. Specific RU Representations	9
14. PSI CV Update procedure.....	10
15. Contributors	10
16. Intellectual Property Statement	11
17. Copyright Notice	11
18. References.....	11

1. Introduction

The structure and management process of the PSI consists of interacting formal “Working Groups” (WGs). Each WG focuses on a specific proteomics technology and is charged with creating, editing, maintaining and obsolescing recommendations for PSI standards. An important subset of such standards are Controlled Vocabularies (CVs), which are associated with a specific WG charter [1].

The MSI is governed by an Oversight Committee on ‘Reporting Standards’ and operates via five WGs addressing different aspects of the standardization process. Three WGs focus on establishing minimal information consensus checklists, and the other two develop the semantic and syntactic solutions required to capture and transmit the information requested in those minimal information checklists.

2. Scope of the document

This document addresses the design principles for the systematic development of CVs, supporting the following tasks:

- T1. *Collecting terms* required to support the minimal information reporting guidelines and the data transfer format for the associated technology [1].
- T2. *Accurately representing terms*, reflecting the existing terminology of the covered technology or target domain.
- T3. *Defining terms via* unambiguous definition of the meaning of a term in the context of covered technology or the target domain
- T4. Providing a community defined and accepted lexicon for a particular working group and ultimately the covered technology or target domain.

The CV MUST be:

- a) *Intelligible*, to a domain expert and to the target community.
- b) *Formalized*, in a sufficiently granular way to facilitate efficient searching, referencing and application of terms.

The following sections present the conventions for complete CVs and for each term of the CVs, which MUST be adhered to when developing CVs under the PSI/MSI umbrella. Defining conventions and the application of these conventions will provide a consistently aligned platform for the individual WGs to develop robust CVs and facilitate the consistent interpretation and interoperability of the CVs by the scientific community.

3. Notational Conventions

In this document the key words “MUST,” “MUST NOT,” “REQUIRED,” “SHALL,” “SHALL NOT,” “SHOULD,” “SHOULD NOT,” “RECOMMENDED,” “MAY,” and “OPTIONAL” are to be interpreted as described in RFC-2119 [11].

Various communities define controlled vocabularies using alternative descriptions [7,8], resulting in multiple definitions without widespread community acceptance. Therefore in the context of this document we propose the following definition:

- A controlled vocabulary (CV) is a set of terms and associated definitions that are controlled through a standardization authority and recommended for general use within an appropriate community. A CV is a representational artifact (RA) [2,7], consisting of terms as the main representational unit (RU). The relations among CV terms are discussed in Section 11. The CV **MUST** be intelligible to the experts in the target domain and **SHOULD** support the tasks outlined in Section 2.
- Each representational unit (RU) **MUST** have an `identifier`, a `label`, a `definition`, a `source` and **CAN** also have `synonym` or `comment`. Moreover each RU **MUST** have at least one relationships to other RU's within a representational artifact (RA) [2,7]. In this document we use `Courier` font to represent RU and RA attributes.

Therefore in this document we use ‘representational unit (RU)’ only as a synonym for ‘CV term’ and refer to other metadata as ‘attribute’ or ‘relation’.

4. CV metadata

Each CV delivered by a workgroup **MUST** have the following minimal set of descriptive and administrative metadata on the RA level (for the whole CV) and on the RU level (for each term). These **SHOULD** be captured in a sufficiently granular and formal way to be tractable and facilitate users of the CVs.

4.1 RA attributes (metadata for the whole CV)

The labels for the following CV metadata are derived from the OBO library [3] and the Dublin Core Metadata Initiative, (<http://dublincore.org>):

- `namespace`: The namespace **SHOULD** be a token which reflects the coverage of the RA. The namespace and the namespace acronym must be unique within the OBO library [3]. The namespace acronym is used as the full CV identifier and as prefix for the RU Identifiers e.g. the acronym `sep` for ‘sample separation’
- `release date`: A point in time associated with an event in the lifecycle of the resource. This is the date of the release of the file.
- `version`: The described resource has a version, edition, or adaptation, namely, the referenced resource. The version number of the released CV file, e.g. `psi-mi25.obo,v1.28`
- `coverage`: The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant. The technology or the target domain represented by the CV, e.g. the coverage for `sep.obo` CV is ‘sample separation procedures’.
- `creator`: An agent primarily responsible for making the resource, for example the name of the person responsible for editing the CV of a working group.

- **publisher**: An agent responsible for making the resource available, for example one or more institutions heading the working group.

4.2 RU attributes (metadata for each CV term)

Each RU or CV term **MUST** be specified with the following minimal set of attributes to describe editorial- and authoring-, status- and versioning-, as well as provenance data:

- **identifier**: The token used to uniquely identify a RU.
- **label**: The human-readable token assigned to the RU.
- **definition**: A natural language statement of the meaning and essential nature of the entity represented by the RU.
- **source**: A reference to a resource from which the present resource is derived; for example the citation source of an RU or its attributes

Where applicable, the following attributes provide additional information about a term:

- **synonym**: An alternative term which serves as a textual or symbolic substitute for the RU.
- **comment**: Additional information about the RU or its application.

5. Identifier

An identifier according to the Dublin Core (<http://dublincore.org>) is an unambiguous reference to the resource within a given context.

5.1 CV Namespace

A CV **MUST** have a namespace to provide a unique and abstract context for its RU's. The working group **MUST** choose a namespace and a namespace acronym which are unique within the OBO Library [3]. The namespace **SHOULD** be referred to through the namespace acronym (or abbreviation) which **SHOULD** consist of the two or more character acronym of the working group itself. In special cases, such as where several working groups develop a single CV collaboratively, the namespace abbreviation **SHOULD** represent the scope of the CV, as far as possible.

When a WG formally releases the first version of its CVs the WG is **REQUIRED** to register a namespace at the OBO library web site [3]. To register the namespace email the OBO webmaster or the obo-discuss@lists.sourceforge.net mailing list, including its administrative RA metadata, such as source institution, contact person, stable location of the CV file and its format (see the full list of details to provide at http://www.geneontology.org/doc/GO.xrf_abbs_spec).

5.2 RU identifier

A RU **MUST** have a unique identifier. The system of identifiers for the CVs consists of two parts. Part one **SHOULD** be the official 'namespace abbreviation' XXX (e.g. 'sep' for Sample processing and separation techniques). The second part corresponds to a numeric accession number of any number of digits. Therefore, the identifier pattern **SHOULD** be XXX:000000 using the colon as separator as recommended in the OBO format [3] (e.g. sep:00041).

6. RU Label Naming Conventions

The following recommendations are derived from the naming conventions developed by the Metabolomics Standards Initiative (MSI) [2] and from the OBO foundry recommendations [3].

A RU Label SHOULD be of minimal length, easy to remember and as self-explanatory as the pragmatic compromise allows.

A Label SHOULD also allow the formation of derivatives and:

- SHOULD be linguistically correct (i.e. they SHOULD conform to the rules of the language in question).
- SHOULD be precise and motivated (i.e. they SHOULD reflect as far as possible the characteristics which are given in the definition and use well accepted terms).
- MUST be concise in capturing the meaning of what the labelled RU represents.

6.1. Character set and word separators

A RU Label SHOULD consist mainly of alphabetical characters and numbers. A white-space character SHOULD be used as the word separator.

Special characters SHOULD be avoided where possible. For instance, underscore, accents, sub- or superscript characters and character-combinations that may have a special meaning in regular expressions or programming languages SHOULD NOT be used.

The hyphen SHOULD be avoided as a word-separator although this does not prevent its use when required by the community for specific representations.

6.2. Spelling and linguistic convention

Language: When there are differences in spelling between British English and American English, the American form SHOULD be used, e.g. “polymerizing”, “signaling” rather than “polymerising”, “signalling”.

Positive: A negative RU Label such as “non-mammal” or “non-membrane” SHOULD be avoided. There are some valid exceptions to this suggestion e.g. an “ex-vivo” role.

Capitalization: An RU Label SHOULD be written in lower case letters throughout except for acronyms (e.g. DNA) which are capitalised (if their use in term names cannot be avoided) and proprietary names, which are written as such.

Tense: An RU Label SHOULD NOT use the tense form. Rather use noun form as far as possible or the present tense when necessary, e.g. “measurement” SHOULD be used instead of “to be measured”, “measuring”, “measurement taken” consistently throughout the CV.

Singular nouns and plurals: An RU Label SHOULD be in the singular form throughout. An RU Label MUST always be a singular noun. If the plural is required, consider using “protocol collection” instead of “protocols”.

Acronyms and Abbreviations: Abbreviations in an RU Label SHOULD be avoided and acronyms resolved. An RU Label SHOULD be explicit, e.g. “number of residues” SHOULD be used instead of the unintuitive “n res”. Acronyms SHOULD be included in the synonyms list and resolved within the definition. When an acronym, however, is commonly used, for example “laser”, “DNA”, it can be used as an RU Label, while its resolved Label SHOULD be listed in the synonym list. Abbreviations which can have other meanings SHOULD NOT be allowed (‘chronic olfactory lung disorder’ SHOULD NOT be abbreviated: cold).

Affixes (prefix, suffix, infix and circumfix): The word-stem SHOULD be used and affixes to names SHOULD be avoided where possible or at least be used consistently. Since each term 'A' implicitly *means* 'the RU A', either prefixes or postfixes involving "RU" SHOULD NOT be used. The same applies to suffixes like "entity" and "type".

Avoid linguistic ellipses: Avoid ellipses and apocope. These are rhetorical figures of speech involving the omission of a word or words required by strict grammatical rules but not by sense. The missing words are implied by the context in human language. For instance, 'HIV' can be an ellipse standing for "HIV Virus", or "HIV Disease"

Registered Product and Company names: Proprietary names SHOULD be captured as they are. This is an exception to the naming conventions.

"Taboo" words and Characters: The use of database or ontology language keywords, for example Booleans or the words "Model", "Class", "Term", "KIF", "Method", "Property", "Relationship", "Clips", "OBO" and "OWL" and xml style tags or characters designating tags or regular expressions SHOULD be avoided. The use of the terms such "experiment" "method" "technique", "instructions" are ambiguous and have many meanings across science as well as across proteomics technologies. Therefore it is RECOMMENDED that, a series of events or actions used in proteomics SHOULD be represented as a single or collection of atomic "Protocols" rather than using the terms above.

7. RU definitions

The `definition` attribute of a RU SHOULD provide the context and meaning of the term in a way that eases its interpretation [2]. As with term names, avoid using unresolved abbreviations. Definitions SHOULD be as brief as possible, but as complex as necessary. They SHOULD begin with an upper-case letter, can consist of more than one sentence if necessary and always end with a period (full stop). Examples MAY be added into a definition typing the string 'EXAMPLE:' before the appropriate part of the definition.

Definitions SHOULD start in the following the Aristotelian way:

"A [term described] is a [parent term], which/that [most relevant intrinsic properties not already mentioned on ancestor levels (attributes and relations to other terms)]".

For example:

'**image acquisition device**' sep:00096 "An image acquisition device is a device which captures a digitised image of an object."

- '**camera**' sep:00099 "A camera is an image acquisition device which is used for taking photographs (usually consisting of a lightproof box with a lens at one end and light-sensitive film at the other)."
- '**scanner**' sep:00100 "A scanner is an image acquisition device that generates a digital representation of an image for data input to a computer."

Words like 'it' SHOULD NOT be used. But when they are necessary, make sure they only refer to the described RU. General words like "generally", "often", "in most cases" SHOULD be avoided to enforce a more concise definition.

7.1 General rules for creating sound normalized definitions

1. A RU MUST only have one natural language definition.

2. Definitions SHOULD use simple, clear, easy to understand words that are meaningful to the users.
3. The definition SHOULD be written at the same level of specificity as the RU itself.
4. Definitions SHOULD define RU and their essence and not the words used to refer to them, so avoid terms such as 'term', 'descriptor', 'name', etc.. E.g. the definition of 'eye' is 'organ of sight', not 'is the name of an organ of sight', nor 'term or concept describing an organ of sight'.
5. Each definition SHOULD reflect the position in the hierarchy of the defined representational unit.
6. The definitions SHOULD explain what are characteristics (or relationships) that distinguish a term from the others (the parent term and siblings).
7. A definition SHOULD be positive and not negative. Definitions such 'all animals that are not a mammal' or 'all non-membrane proteins' SHOULD be avoided.
8. The definition SHOULD be correct in all possible contexts the RU is used, so that the RU and all its synonyms are intersubstitutable with its definition in such a way, that the result is both grammatically correct and truth preserving.
9. The definition SHOULD avoid circularity like these:
 - a) An A is an A which is B (person = person with identity documents)
 - b) An A is the B of an A (heptolysis = the causes of heptolysis)

8. Source of RU Definitions

A Source is "a reference to a resource from which the term, its label, definition and placement in the hierarchy are derived"(<http://dublincore.org>) Every RU Definition MUST have a documented source; i.e. the provenance of the information where the definition originated from. A RU Definition Source SHOULD consist of three parts:

- The `Source_Label`: the name of the source
- The `Identifier_in_source` used within the source to represent the term or any document relative to the term. This identifier SHOULD be a URI, When it is not possible to specify a URI, then an internal identifier of the source can be used instead. If no internal identifier exists then use the value "unidentified".
- The `Identifier_of_source`. This SHOULD be the identifier of the source in PSI-Common CV. Until the PSI-Common CV is officially released this SHOULD be a free text description.

Examples are given below for specific sources:

Database source: For example, the definition source derived from the protein GTR1_MOUSE (<http://www.expasy.org/sprot/userman.html> - AC line **P17809**) in the Uniprot database SHOULD be represented as "uniprot:<http://www.expasy.org/uniprot/P17809>: [id of uniprot in PSI common]"

The definition source of the journal article identified by the token PMID:10350628, in the PubMed database, SHOULD be represented as "pubmed:<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=search&term=10350628>: [id of pubmed in PSI common]"

Book source: If the RU Definition Source is a book, use the ISBN database. For example, any RU derived from The dictionary of Cell & Molecular Biology SHOULD have "isbndb:<http://isbndb.com/search-all.html?kw=0124325653>: [id of isbndb in PSI common]" Hyphens SHOULD be removed from the ISBN.

Digital Publication source: If the Term Definition Source is a digital publication then use the appropriate identifier such as the digital object identifier (DOI) “DOI System resolver: <http://dx.doi.org/10.1108/00251740410510190>: [id of DOI System resolver in PSI common]”,. For the term “proteomics” taken from Wikipedia the Term definition source would be “Wikipedia:<http://en.wikipedia.org/wiki/Proteomics>: [id for Wikipedia in PSI common]”

PSI Working Group source: When the definition of a particular term does not exist or is not adequate to describe the meaning within the proteomics domain, the relevant PSI WG is responsible for creating the definition. As a result the source of this definition SHOULD be identified using the official ‘namespace abbreviation’, the 2 or more character acronym of the working group itself. For the term “image acquisition” created by the PSI Gel WG the Term Definition source SHOULD be “psi-gel:<http://psidev.sourceforge.net/gps/CV/>: [id for PSI-GEL in PSI common]”.

9. RU Synonyms

An RU `synonym` is an optional RU attribute. A synonym is “A word or an expression that serves as a textual or symbolic substitute for another” [10] and can be used interchangeably in all contexts. The number of synonyms for a RU is not limited. Acronyms and symbols are synonymous with the full name.

A synonym SHOULD have a documented source; i.e. the provenance of the information where the definition originated from and MUST be represented in the same manner as a RU definition source (Section 8).

10. RU Comments

An RU `comment` is an optional RU attribute. It MAY be a free text annotation providing additional discourse about the term, not presented within the RA or by the compulsory RU attributes. It SHOULD NOT be used to duplicate information presented within the compulsory attributes of the RU.

11. Relations between RU's

As the PSI CV will be developed under the OBO umbrella [3], the relations created between terms MUST ascribe to the definitions and formal requirements provided in the OBO Relations Ontology (RO) paper [7], as the relations ‘is_a’ and ‘part_of’.

12. Obsolescing RU's

Following GO editing guidelines (at <http://www.geneontology.org/GO.usage.shtml>) a RU which is no longer used MUST NOT be deleted, but tagged as 'obsolete' [6]. A unique identifier MUST NOT be deleted once used. IDs and the corresponding terms MUST be conserved at all times so that, even if a term is defunct or has a new ID, someone searching using the old ID can find it.

A term can become obsolete when it is merged, split, replaced or deprecated, but a term MUST NOT be made obsolete due to changes in wording that do not alter the meaning of the term. When a term's definition changes meaning, the term SHOULD also be assigned a new ID, and the old term considered obsolete.

To request the obsolescence of a RU, the CV update procedure **SHOULD** be followed (Section 13.). When you make a term obsolete, insert the word 'OBSOLETE' at the beginning of the term definition and add a comment that explains why the term has become obsolete and suggests alternative terms for annotators to use. Use the following syntax for the reason for obsolescence

comment: This term was made obsolete because [reason].

The [reason] **SHOULD** indicate the cause leading to the obsolescence of a RU. The reason **SHOULD** reported using one of the following keywords '**merge**', '**split**', '**replacement**' or '**simple deprecation**'. The reason '**merge**' **SHOULD** be used when two or more RUs result to be semantically redundant, one RU is chose to remain the representative one and the others **SHOULD** become obsolete. In reverse, when an ambiguous RU becomes obsolete and is replaced by more than one RU the reason '**split**' **SHOULD** be reported. The keyword '**replacement**' **SHOULD** be used in cases where a term name and definition have changed substantially and lead to the creation of a new RU and the obsolescence of the original one. The reason '**simple deprecation**' indicates a RU **SHOULD NOT** be used and there are no remapping suggestions within the same RA.

12.1 Alternative terms for obsolete RU's

Use the following syntax to suggest alternative terms within a CV to be used instead of, or as a replacement for, obsolete terms:

comment: Alternative term [RU ID]

Use the following syntax to suggest alternative terms within an alternative resource to be used instead of, or as a replacement for, obsolete terms:

comment: Alternative term [Alternative resource] [Alternative resource ID]

12.2. Restoring obsolete terms

If you need to reinstate an obsolete term back into the CV, use the following:

comment: This term was reinstated from obsolete.

13. Specific RU Representations

Developing CVs is a process of collecting, and if necessary defining RU's. Every effort **MUST** be made to adopt and re-use existing ontologies or CVs where they exist, to avoid "re-inventing the wheel". As stated by OBO "we would strive for community acceptance of a single ontology for one domain, rather than encouraging rivalry between ontologies". Therefore it is **RECOMENDED** to represent the following concepts as described.

Units of Measure

The CVs "SHOULD NOT" contain any units of measure. It is **RECOMMENDED** to use, and to contribute to the Unit ontology by requesting required terms (http://obo.cvs.sourceforge.net/*checkout*/obo/obo/ontology/phenotype/unit.obo) via the mailing list obo-unit@lists.sourceforge.net.

Chemical Entities

For the representation of Chemical entities it is **RECOMMENDED** to use terms from Chemical Entities of Biological Interest [ChEBI CheBI <http://www.ebi.ac.uk/chebi/>]. ChEBI is also available from the OBO website [3].

Phenotypic quality

For the representation of Phenotypic quality (e.g. age, color, shape etc..) it is RECOMMENDED to use terms from the quality ontology http://obo.cvs.sourceforge.net/*checkout*/obo/obo/ontology/phenotype/quality.obo) and request any missing term via the mailing list obo-phenotype@lists.sourceforge.net.

14. PSI CV Update procedure

CVs MUST be maintained. The addition or obsolescing of terms SHOULD be done at the requests of the WG or the proteomics community.

Procedure for dynamic maintenance of PSI CVs:

- A CV committee consisting of an odd number of members is appointed by the workgroup chair and ontology chair. The members of the CV committee SHOULD be frequent users of the CV.
- The editor of the CV SHOULD be the ontology chair of the workgroup The CV editor MUST be a member of the CV committee.
- A term request SHOULD be submitted via a dedicated issue tracker or mailing list.
- The request is processed by the editor and sent for approval to the other members of the committee. No reply within a defined period is considered as an agreement. In the case of a disagreement, the request is decided by a committee members vote.
- Within five working days after the last email exchange among the committee members, the new term SHOULD be added to the CV by the CV editor.

15. Contributors

Luisa Montecchi-Palazzi
European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge, CB10 1SD, United Kingdom
luisa@ebi.ac.uk

Frank Gibson
School of Computing Science, University of Newcastle upon Tyne,
Newcastle upon Tyne, NE1 7RU United Kingdom
Frank.Gibson@newcastle.ac.uk

Daniel Schober
European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge, CB10 1SD, United Kingdom
schober@ebi.ac.uk

Susanna Sansone
European Bioinformatics Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge, CB10 1SD, United Kingdom
sansone@ebi.ac.uk

The authors gratefully acknowledge the contributions of all the members of the MSI and PSI who actively participate to this document revision. We also want to thank Barry Smith and Wacław Kusnierczyk for their careful review of this document and Gilberto Fragoso, Trish Wetzel, Matthew Pocock and Micheal Ashburner for their stimulating suggestions. Phillip Lord for review of the original draft manuscript. Finally we are grateful to Amelia Ireland from the OBO team for her the kind availability to answer all our questions and requests.

16. Intellectual Property Statement

The PSI takes no position regarding the validity or scope of any intellectual property or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; neither does it represent that it has made any effort to identify any such rights. Copies of claims of rights made available for publication and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the PSI Chair.

The PSI invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights which may cover technology that may be required to practice this recommendation. Please address the information to the PSI Chair (see contacts information at PSI website).

17. Copyright Notice

© 2007 [Proteomics Standards Initiative / Metabolomics Standards Initiative]

This is a recommendation document distributed under the terms of the complete Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.5>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

18. References

- [1] Julian R, Hermjakob H, "A Proposed Management Structure and Document Process for the HUPO PSI" April 2006
- [2] Metabolomics Society Ontology Working Group, Naming Conventions Strawman document (NCstrawman.doc), <http://msi-ontology.sourceforge.net/recommendations/>
- [3] Open Biomedical Ontology (OBO): <http://obo.sourceforge.net/>
- [4] ISO standards (<http://www.iso.org/iso/en/ISOOnline.frontpage>)
ISO 704:2000 Terminology work – Principles and methods
- [5] ISO standards (<http://www.iso.org/iso/en/ISOOnline.frontpage>)
ISO 1087-1:2000 Terminology work – Vocabulary – Part 1: Theory and application
- [6] The Gene Ontology Consortium, Gene Ontology: tool for the unification of biology. (2000) *Nature Genet.* 25: 25-29.
- [7] Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical Ontologies. *Genome Biol.* 2005; 6(5): R46

- [8] Metamodel.com “What are the differences between a vocabulary, a taxonomy, a thesaurus, an ontology, and a meta-model?”
<http://www.metamodel.com/article.php?story=20030115211223271>
- [9] Wikipedia. “Controlled Vocabulary” http://en.wikipedia.org/wiki/Controlled_vocabulary
- [10] Dictionary definition of **synonym** on Answers.com. The American Heritage® Dictionary of the English Language, Fourth Edition Copyright © 2004 by Houghton Mifflin Company. Published by Houghton Mifflin Company.
- [11] S. Bradner, *Key words for use in RFCs to Indicate Requirement Levels*, Internet Engineering Task Force, RFC 2119, <http://www.ietf.org/rfc/rfc2119.txt>, March 1997.