



Information Management Tool:  
Guide to the Development and  
Maintenance of Controlled Vocabularies in  
the Government of Canada  
2nd Edition



Chief Information Officer Branch  
Information Management Division

July 2013

## Version Control

Date	Action
2012-11-21	Final draft presented to Metadata Working Group (Patricia Redmond)
2012-12-19	Endorsed by Metadata Working Group (Patricia Redmond)
2013-03-15	Translation completed (Marie-Claude Côté)
2013-07-17	Edits completed (Marie-Claude Côté)

*Guide to the Development and Maintenance of Controlled Vocabularies in the Government of Canada, 2<sup>nd</sup> Edition*

Prepared by the Controlled Vocabulary Guide Task Group of the Government of Canada Metadata Working Group

© Her Majesty the Queen in Right of Canada, represented by the President of Treasury Board, 2012

The approved version of this document will be available on the [Treasury Board of Canada Secretariat website \(www.tbs-sct.gc.ca\)](http://www.tbs-sct.gc.ca).

This document supersedes the *1st edition: Guide to the Development and Maintenance of Controlled Vocabularies in the Government of Canada, July 2005*<sup>1</sup>.

This document is available in alternate formats on request.

Please direct enquiries about this guideline to your department's area of responsibility for IM.

This guide will continue to evolve over time, based on feedback and suggestions from related communities of interest.

Has this document been helpful to you? Do you have suggestions to improve it? Please send all suggestions or enquiries to the following:

Information Management Division  
Chief Information Officer Branch

Treasury Board of Canada Secretariat  
8th Floor, 270 Albert Street  
Ottawa, ON K1A 0R5

**Email:** [im-gi@tbs-sct.gc.ca](mailto:im-gi@tbs-sct.gc.ca)

**Toll-free:** 1-877-636-0656

**Fax:** 613-946-9342

**TTY:** 613-957-9090 (Treasury Board of Canada Secretariat)

---

<sup>1</sup> [http://www.collectionscanada.gc.ca/webarchives/20071207085552/http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/cvsg-sgvc/docs/2005/vocab/vocab00\\_e.asp](http://www.collectionscanada.gc.ca/webarchives/20071207085552/http://www.tbs-sct.gc.ca/im-gi/mwg-gtm/cvsg-sgvc/docs/2005/vocab/vocab00_e.asp)

## Acknowledgements

This guide is a project of the Controlled Vocabularies Guide Task Group of the Metadata Working Group (MWG). It draws on the expertise of information management professionals in several disciplines and government departments who pooled their experience on the issues and best practices involved in the application, maintenance and development of controlled vocabularies for the Government of Canada.

For their support in the creation of this guide, the Task Group thanks:

Name	Department
Noris Birt	National Research Council Canada
Minda Bojin	Treasury Board of Canada Secretariat
Marie-Claude Côté	Treasury Board of Canada Secretariat
Alison Hayman	Statistics Canada
Julia Lipinska	Health Canada
Stephanie Osley	Agriculture and Agri-Food Canada
Patricia Redmond (editor)	Agriculture and Agri-Food Canada
Cecil Somerton	Fisheries and Oceans Canada
Paula Tozer	Environment Canada

# Table of Contents

Chapter 1: Introduction.....	6
1.1 Government of Canada Context .....	6
1.2 What Is a Controlled Vocabulary? .....	6
1.3 Why Are Controlled Vocabularies Useful? .....	7
1.4 Controlled Vocabularies for GC Metadata Elements .....	8
Chapter 2: Structure of Controlled Vocabularies .....	9
2.1 Vocabulary Control Forms .....	10
2.1.1 Grammatical Forms.....	10
2.1.2 Singular and Plural Forms.....	11
2.1.3 Punctuation and Capitalization.....	12
2.1.4 Abbreviations, Initialisms, and Acronyms .....	12
2.2 Selection of Terms.....	12
2.3 Justification of Terms.....	13
2.4 Controlled Vocabularies with Preferred Terms .....	14
2.5 Controlled Vocabularies with Related Terms.....	14
2.6 Controlled Vocabularies with Term Definitions .....	15
2.7 Equivalence Relationships.....	16
2.7.1 Linguistic .....	16
2.7.2 Synonymy .....	16
2.7.3 Lexical Variants.....	17
2.7.4 Compound Terms .....	17
2.7.5 Quasi-synonyms .....	18
2.7.6 Generic Posting.....	19
2.8 Associative Relationships .....	19
2.9 Hierarchical Relationships .....	20
Chapter 3: Assessing Existing Controlled Vocabularies.....	23
3.1 Identifying and Examining the Existing Problem.....	23
3.2 Scoping the Solution.....	23
3.3 Selecting Available Controlled Vocabularies .....	24
3.4 Evaluating a Controlled Vocabulary for Use .....	25
Chapter 4: Design and Construction of New Controlled Vocabularies .....	27
4.1 Scope and Use of the Information Domain .....	27
4.2 Determine Function, Size and Structure.....	28
4.3 Concept Representation.....	28

Chapter 5: Management and Maintenance of Controlled Vocabularies .....	30
5.1 Governance Structure .....	30
5.2 Editorial Policy .....	30
5.3 Indexing Policy .....	31
5.4 Maintenance Policy .....	31
5.5 Publication Policy .....	32
5.6 Technology .....	33
Appendix A: Glossary .....	35
Appendix B: Selected Bibliography .....	43
Appendix C: Guidance and Standards .....	45

# Chapter 1: Introduction

## 1.1 Government of Canada Context

This guide is intended for the use of information professionals within the Government of Canada (GC) who are faced with the challenges of extending or mapping existing controlled vocabularies. The emphasis is on controlled vocabularies that will provide terms for GC metadata elements. This guide is limited predominantly to existing controlled vocabularies. A brief overview of the construction and creation of a new controlled vocabulary is provided for information purposes. To describe the process in detail, however, is out of scope because it requires professional expertise and resource materials specific to creating controlled vocabularies.

This guide addresses controlled vocabularies in the context of requirements for creating and maintaining metadata and value domains as mandated by the GC *Standard on Metadata* (<http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909>). Controlled vocabularies are primarily used within the Government of Canada for information resources in conjunction with metadata.

The *Standard on Metadata* (<http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909>), states that: “along with standardized metadata, consistent use of value domains supports the exchange of information resources within and across systems. Value domains include, but are not limited to, controlled vocabularies, taxonomies, ontologies, thesauri, reference data, authority lists, naming conventions, and value string instructions.”

## 1.2 What Is a Controlled Vocabulary?

The GC *Standard on Metadata* definition of a controlled vocabulary is:

“A list of standardized terminology, words, or phrases, used for indexing or content analysis and information retrieval, usually in a defined information domain.”

Controlled vocabularies are characterized by a consistent format, syntax and grammar and may include synonyms and cross-references. Controlled vocabularies can apply to many different concepts, including subjects of resources, their formats, types, or the audiences for which the resources are intended.

All controlled vocabularies have a predetermined, explicit, and coherent structure. Terms within a controlled vocabulary are usually organized and classified according to an existing classification scheme, this scheme may be a recognized classification scheme, such as the Dewey Decimal Classification System or a scheme devised by the rules set out in the editorial policy of an organization. Controlled vocabularies are used by many different types of users. Indexers, content analysts, or metadata creators can use controlled vocabularies as analytical or descriptive tools. Information users or information managers use it for information organization or retrieval. Information architects create taxonomies to map content resources.

### 1.3 Why Are Controlled Vocabularies Useful?

Controlled vocabularies provide a way to organize information resources. They create a standardized and authoritative way of structuring information. Here are some of the key functions that a controlled vocabulary can render:

- Findability  
*Allows users to retrieve information efficiently through search or navigation*
- Relationships  
*Establishes the relationships among terms and whether a term is a related, broader, or narrower term*
- Disambiguity  
*Eliminates ambiguity between terms and brings together similar terms*
- Interoperability  
*Facilitates the movement and sharing of specific information resources within and between systems and destinations*
- Consistency  
*Controls synonyms, homonyms, common misspellings, pseudonyms, name changes, and connects abbreviations to the full word and variant spelling of terms*
- Concept Mapping  
*Creates information models for information architecture and business analysis*

It is important to note that unless there is a technology designed to make appropriate use of the controlled vocabulary, it is impossible to take advantage of the vocabulary. For indexing and findability to be functional as well as useful for users, search engines must be specifically configured to search for controlled vocabulary terms.

By combining controlled vocabularies with sophisticated search algorithms, complex searches can be performed allowing deeper access to content with greater precision.

## 1.4 Controlled Vocabularies for GC Metadata Elements

The GC departments and agencies have registered numerous controlled vocabularies, such as, subject, format, type, etc. For example, the *Standard on Metadata* (<http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909>) recommends that the *Government of Canada Core Subject Thesaurus* (CST) (<http://www.thesaurus.gc.ca>) be used as the default controlled vocabulary to describe the subject(s) of GC Web resources. This means that, in the absence of any other appropriate registered subject vocabulary, the CST must be used as a source of vocabulary for subject.

As prescribed in the *Standard on Metadata* (<http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909>), Library and Archives Canada (LAC) is responsible for maintaining the CST and ensuring that it remains publicly available.

LAC also ensures the registration of in use controlled vocabularies within the GC or that have been approved by interdepartmental or regulatory agencies designated by TBS. LAC maintains a registry of authorized controlled vocabulary schemes (<http://www.collectionscanada.gc.ca/government/controlled-vocabularies/007004-1000-e.html>).



## Chapter 2: Structure of Controlled Vocabularies

Folksonomies and social tagging are sometimes confused with controlled vocabularies, however, they are *not* controlled vocabularies. Folksonomies have no standardization or common vocabulary employed, it is a decentralized practice among individuals who create keywords for digital resources.

All controlled vocabularies have authoritative terms, meaning that preferred terms and non-preferred terms have been identified. Preferred terms are also referred to as the “top term”.

**Flat file:** It is the simplest form of a controlled vocabulary. It does not have any structure or relationship between the terms within the list but maintains ambiguity control with preferred terms.

**Synonym ring:** It is slightly more complex than an authority file. It has no preferred term, which mean all terms are equal and therefore point to each other.

**Taxonomy:** It has more complexity than a synonym ring and includes ambiguity and synonym control and hierarchical relationships between terms (BT, NT, RT). A taxonomy typically, has a hierarchical structure that includes parent/broader terms (BT), child/narrower terms (NT), and related terms (RT). Taxonomies are often displayed as a tree structure.

**Thesaurus:** It is the most complex of all controlled vocabularies. It includes all the complexity of a hierarchical controlled vocabulary as well as associative (*see also*) and equivalent (*use/used for*) relationships, definitions, and scope notes. The nature of a thesaurus can be polyhierarchical which allows for it to expand beyond content navigation and term standardization into an enhanced information retrieval tool.

**Ontology:** It is similar to a taxonomy with structure and specific types of relationships between terms. In an ontology the types of relationships are greater in number and more specific in their function. Relationships could include, for example, *located in* to relate an organization to a place, *produced by* to relate a company and its product , etc. It is a framework that defines concepts and the relationships between them and the nature of the relationships. All details of concepts, properties, and constraints within a knowledge domain are documented. An ontology is usually in a form that is both machine readable and interpretable. It could be encoded to a scheme such as, XML,

RDF, or OWL, which are meta-ontologies; and their instances are Semantic Web ontologies. Ontologies can make inferred relationships possible.

**Linked Data & Semantic Web:** There are many semantic web technologies that can utilize linked data. Linked data is a way of publishing data on the semantic web that encourages reuse, reduces redundancy, and allows for real and potential inter-connectedness. This will enable network effect that will add value to any linked data.

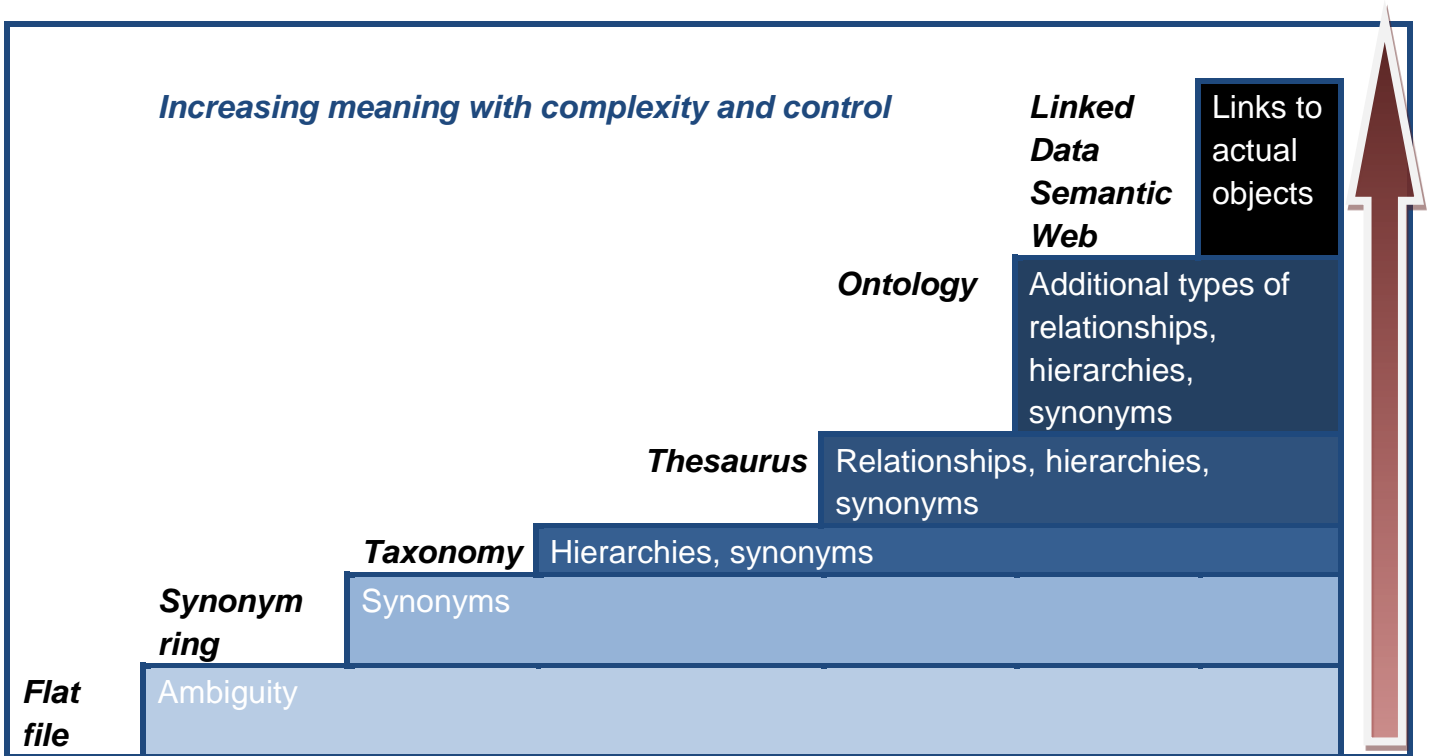


Figure 1. Building complexity

## 2.1 Vocabulary Control Forms

Natural language contains a multitude of words and phrases that designate particular concepts. These include homographs, homonyms, synonyms, quasi-synonyms, and lexical variants, among others, and all of them may cause problems of ambiguity and/or redundancy. By selecting a single term from among many possibilities, controlled vocabularies reduce these semantic problems to a minimum. The following are some examples of vocabulary control:

### 2.1.1 Grammatical Forms

It is important to have a high degree of consistency. It is necessary to establish the grammatical forms of words or phrases that are used as preferred and non-preferred terms. This will establish a uniform and consistent form for terms and their display.

- Nouns and noun phrases  
*Used for indexing terms; noun phrases are commonly used as modifiers*
- Adjectives  
*Used only in compound indexing terms*
- Adverbs  
*Used only in compound indexing terms*
- Verbs  
*Usually excluded from subject vocabularies; probably required to some degree in functional vocabularies*
- Initial articles  
*Initial articles should be avoided if possible*

### **2.1.2 Singular and Plural Forms**

It is necessary to have a standardized approach that establishes the terms with a high degree of consistency in either their singular or plural form; either form can be used in particular circumstances, by common convention, English terms generally appear in their plural form. In contrast, French terms by convention are generally established in their singular form, but there are also many exceptions in French vocabularies that are determined by the meaning established for the term and usage. French terms by convention are generally established in their masculine form where there is a possibility to have a masculine and a feminine form.

In addition, the following elements, in particular, should be considered.

- Concrete entities  
*In English, by convention, if something is countable (e.g. automobiles), use the plural form. In French, use the singular form (e.g. automobile). For non-count nouns and collective nouns, use the singular form in English (e.g. sand) and French (e.g. sable).*
- Abstract concepts  
*In English and French, by convention, use the singular form.*
- Unique entities  
*In English and French, by convention, use the singular form.*
- Co-existence of singulars and plurals

*When the singular and plural forms have different meanings, a qualifying word or phrase is added in parentheses.*

- Standardization

*Standards for spelling, transliteration, romanization, etc. should be established in editorial policy, if applicable.*

### **2.1.3 Punctuation and Capitalization**

It is important to establish a standardized approach to punctuation and capitalization practices.

- Parentheses

*Parentheses are used only to enclose qualifiers.*

- Hyphens

*Avoid the use of hyphens wherever possible.*

- Capitalization

*Capitalization is used only for initial letters of proper names, by convention for the initial letters of indexing terms, and for acronyms.*

### **2.1.4 Abbreviations, Initialisms, and Acronyms**

An acronym or an abbreviation can become universally known to the point where the abbreviated form is the most familiar and the full form is treated as the non-preferred term.

For example, the term “BSE” can be used as opposed to “Bovine Spongiform Encephalopathy”

The selection will be influenced by the audience. Consistent application of the selection is necessary.

## **2.2 Selection of Terms**

You must consider and decide whether or not the following will be included:

- Loan words
- Neologisms, slang terms, and jargon
- Common names versus trade names

- Trade marks
- Popular names and scientific names
- Place names (the GC has registered geospatial vocabularies)
- Proper names of institutions, persons, etc. (identifiers)

## 2.3 Justification of Terms

Literary warrant is usually the most reliable basis for term inclusion, but user warrant and organizational warrant may be good authorities for terminologies that will be easily recognized and understood by members of organizations or researchers. Both may be a good basis for establishing synonyms.

- Literary warrant refers to a justification for selecting terms based on a significant frequency of occurrence of those terms in the information resources to be indexed. Literary warrant ensures that resources that could be indexed using the terms in the vocabulary actually exist. This guarantees that unnecessary terms are not included in the vocabulary. It also ensures that the form of term selected as the preferred term will be the most commonly used in information resources.
- User warrant refers to a justification for selecting terms based on words or phrases employed by users of information resources for information retrieval or information management. Evidence of such usage may be derived from search engine logs or interviews. User warrant ensures that the language of the vocabulary matches the language of the user community. For example, search logs can indicate that users search using terms such as 'employment,' 'jobs,' but do not search using terms such as 'labour,' 'career opportunities' or 'staffing'. Knowing this will inform the selection of controlled vocabulary terms and the possible use of non-preferred terms.
- Organizational warrant refers to a justification for selecting terms based upon the business requirements and the business language used by an organization. Determining organizational warrant requires identifying the forms that are preferred by the organization that will use the controlled vocabulary. Organizational warrant ensures that the language of the vocabulary matches the needs and priorities of the organization. For example, an organization which works with alternative crops may need to provide narrower terms for the different alternative crops such as buckwheat and lentils which fall under this category. Other organizations do not have a justification for such granularity.

## 2.4 Controlled Vocabularies with Preferred Terms

In natural language, a given concept is often represented or referred to by means of a wide range of words or phrases that express or imply a variety of contexts, shades of meaning, or application. Controlled vocabularies identify and restrict the many possibilities to a single or very limited meaning for the purposes of indexing, information retrieval, or information management. To ensure consistency in the description of concepts, some controlled vocabularies guide indexers and searchers from a set of possible terms representing a concept to the designated or preferred term for that concept. Consequently, all resources about, or pertinent to, that particular concept, within a body of information resources, can be indexed using this single representative term.

UF is the abbreviation of “Used for.” This entry indicates that, in the context of the *Core Subject Thesaurus*, the concept of “Atmospheric pollution” should be described using the term “Air pollution.” Those considering searching for “Atmospheric pollution” will be guided to the preferred term.

The following is an example of a preferred term:

Air pollution	UF	Atmospheric pollution
---------------	----	-----------------------

Non-preferred terms are also known as “lead-in” terms. Lead-in terms are synonyms of the preferred term that has been chosen from all terms as the only one authorized to represent the given concept. This is an essential element of “control” in controlled vocabularies.

The following is an example of a non-preferred term:

Atmospheric pollution	USE	Air pollution
-----------------------	-----	---------------

## 2.5 Controlled Vocabularies with Related Terms

Related terms (RT) are terms with conceptual linkages to a given term. Related terms assist indexers and users to change or supplement their indexing or searching strategies respectively.

Example:

Air pollution	RT	Acid rain
---------------	----	-----------

Related terms are always reciprocal, e.g.:

Acid rain	RT	Air pollution
-----------	----	---------------

## 2.6 Controlled Vocabularies with Term Definitions

Controlled vocabularies with more complexity may include the following:

**Parenthetical qualifier:** A word or phrase in parentheses after the first word or words in a term.

The qualifier is used to clarify the meaning of the term. In these cases, the entire text string, including the qualifier, constitutes the indexing term. In some vocabularies, square brackets are used instead of parentheses.

Example:

Acquisitions (Businesses)
---------------------------

**Term definitions:** Explains how to apply the term.

Term definitions are usually prescriptive or restrictive, i.e. they either dictate or restrict the meaning of the term. The definition is not to be used as part of the indexing term.

Example:

abstract	A summary of a document or text
----------	---------------------------------

**Scope notes:** Definitions interpreting the term, particularly in thesauri.

Scope notes are indicative, explanatory, or prescriptive. Scope notes are usually indicated by the abbreviation "SN."

The purpose of a scope note is to:

- Reduce ambiguity
- Increase clarity of meaning
- Indicate specificity
- Indicate restrictions on meaning
- Indicate the range of topics covered

- Provide instructions to indexers

The following is an example of a preferred term with a scope note:

Estates	SN	Investments, money, property or other valuables belonging to a deceased person. NOT to be used in the sense of: Landed property; individually owned piece of land containing a residence, esp. one that is large and maintained by great wealth.
---------	----	--

## 2.7 Equivalence Relationships

The equivalence relationships include synonyms, lexical variants, quasi-synonyms, and factored and unfactored forms of compound terms.

### 2.7.1 Linguistic

In the GC, all published or registered vocabularies must be available in both official languages. Additionally, all GC metadata records must have metadata in the language of the target information resource. Therefore, in most cases, controlled vocabularies must have a field or data element for every preferred term that identifies the equivalent term in the other official language. This linguistic equivalent may not necessarily be a direct translation. Some terms in one language may have more than one equivalent in the other.

Example:

Solar Heating	in French: Chauffage + Energie Solaire
---------------	--

### 2.7.2 Synonymy

Synonyms are different terms for the same concept. Synonymous equivalence relationships occur when a semantic relationship is established between a preferred term and a non-preferred term so that the two terms, so far as indexing is concerned, refer to the same concept.

Types of synonymous relationships include:

- Popular names / scientific names/ acronyms
- Common nouns / trade names
- Standard names / slang



- Terms of different linguistic origin
- Terms originating in different cultures sharing a common language
- Competing names for emerging concepts
- Current / obsolete

### 2.7.3 Lexical Variants

Lexical variants are different terms for the same expression. These variations may stem from spelling or grammatical differences or from abbreviated formats.

- **Variant spellings**

Includes stem variants and irregular plurals.

Examples:

Gipsies/Gypsies
-----------------

Mouse/mice
------------

- **Direct and indirect forms**

Entry under indirect form is required only where the focus of the compound term is not an indexing term. In some thesauri both direct and indirect forms of compound terms are made accessible, not by entry terms in the main alphabetical display, but by means of a separate permuted index. Entry under the indirect form is unnecessary where the two syntactic variants may be described by an algorithm either built into the system or easily constructed by a user at the time of each search.

Example:

educational materials / materials, educational
--

### 2.7.4 Compound Terms

In general, controlled vocabulary terms should be single words representing single concepts. In larger vocabularies designed to index complex information domains, compound terms are common. If compound terms are employed as indexing terms, then an editorial policy that establishes guidelines for the admission of such terms should be created. It can be difficult to know when to factor compound terms into simpler terms, and when it may be better to retain the compound term (also known as the pre-coordinated term). A thesaurus having a majority of single terms is said to have a low pre-coordination level, and one with many two or three word compound terms is

said to have a high pre-coordination level. The pre-coordination level of terms is directly related to the level of specificity of the vocabulary. Post-coordinated terms are a combination of preferred terms. They are only created at the time of searching and are usually complicated and requires manual indexing. Post-coordination is not used as often as pre-coordination. *The Art and Architecture Thesaurus Online* (<http://www.getty.edu/research/tools/vocabularies/aat/index.html>) of The Getty Research Institute has established guidelines for compound terms that are a good starting point for establishing an editorial standard in this regard.

- **Factoring of compound terms**

There are two techniques for factoring compound terms – semantic factoring and syntactical factoring. They re-express compound terms into separate components. Semantic factoring is not recommended due to precision loss.

Semantic factoring example:

Cardiac failure could be re-expressed as Heart + Output + Below + Normal

Syntactical factoring example:

Copper mine splits into two terms: Copper + mine

### 2.7.5 Quasi-synonyms

Also known as near-synonyms, quasi-synonyms are terms whose meanings are generally regarded as different in ordinary usage. However, they are treated as though they are synonyms for indexing purposes. The extent to which terms are considered and treated as quasi-synonyms is dependent on the size of the domain covered by the controlled vocabulary. Quasi-synonyms may include antonyms or represent points on a continuum. Quasi-synonyms include terms that have a significant overlap. One must be careful not to employ this as a means of reducing the size of the vocabulary by grouping together terms that ought to be treated as independent indexing terms, except in marginal subject areas. The statistical frequencies of the individual terms, as seen in their literary warrant, may help to guide the decisions on the inclusion or exclusion of quasi-synonyms.

Examples:

breweries / beverage industry  
sea water / salt water (variant terms)  
height / depth  
literacy / illiteracy (antonyms)  
meteors / meteorites / meteoroids (points on a continuum)

For each set of quasi-synonyms, a designated preferred term should be selected and the other terms associated with it will be recorded as well.

Terms should be treated as quasi-synonyms only in subject areas that are peripheral to the information domain it is representing. If concepts can be identified with sufficient precision to justify their representation as separate terms, they should be individually defined and retained. If two concepts cannot be consistently and reliably differentiated from each other, then a term for one concept should be selected as the preferred term and a USE reference made from the other.

### 2.7.6 Generic Posting

Also known as upward posting, generic postings treats narrower terms as if they are equivalent to, rather than as a species of, their broader terms. The effect is to reduce the size of the vocabulary, but at the same time to retain access via specific terms to the broader terms used to represent them.

Example:

Thermodynamic properties	
UF	Enthalpy
	Entropy
	Free energy
	Heat of absorption
Enthalpy	
USE	Thermodynamic properties

## 2.8 Associative Relationships

Associative relationships are related terms (RT) and have a close or significant semantic relationship but one that is neither hierarchical nor equivalent (synonymous). An associative relationship provides a suggestion to an indexer or searcher to consider terms that are commonly linked in various ways in information resources, fields of knowledge, or in natural language. Two general rules are:

- One of the terms should be strongly implied, according to the frames of reference shared by the users of the thesaurus, whenever the other is employed as an indexing term; and
- One of the terms is a necessary component in any definition or explanation of the other.

Types of Associative Relationships include:

- Same category
- Terms with overlapping meanings (e.g. Ships and Boats)
- One concept is derived from the other
- Different categories
- The whole-part associative relationship (e.g. Harbours—Wharfs)
- A discipline or field of study versus the objects or phenomena studied (e.g. Ornithology—Birds)
- An operation or process versus the agent or instrument (e.g. Photocopying—Photocopier)
- An occupation versus the person in the occupation (e.g. Nursing—Nurse)
- An action versus the product of the action (e.g. Photocopying—Photocopies)
- An action versus its patient (e.g. Food inspection—Food)
- Concepts versus properties of those concepts (e.g. Paint—Colour)
- Concepts versus the origins of those concepts (e.g. Children—Parents)
- Concepts versus causal dependence (e.g. Explosives—Explosions)
- A thing or action versus its counter-agent (e.g. Head injuries—Helmets)
- Raw material versus product (e.g. Iron ore—Steel)
- An action versus an associated property (e.g. Food inspection—Food safety)
- A concept versus its opposite (antonym not treated as a quasi-synonym) (e.g. Imports—Exports)

## 2.9 Hierarchical Relationships

Hierarchical relationships are based on levels of superordination and subordination, where the superordinate term represents a class or a whole and is labelled as the broader term (BT), and the subordinate terms refer to its parts, or narrower aspects of the class (NT).

An example of a hierarchy is:

Air pollution	BT	Pollution
	NT	Smog

Hierarchical relationships are always reciprocal, e.g.:

Smog	BT	Air pollution
Air pollution	NT	Smog

Some controlled vocabularies provide links between terms representing broader and/or narrower concepts. Such hierarchies help users match their specific search needs. Controlled vocabularies that have purely hierarchical structures are frequently identified as taxonomies. Taxonomies that are designed for information management purposes may have multiple hierarchies, i.e. they may be polyhierarchical.

Hierarchical relationships include:

- Generic relationship (“is a ...”):  
*Identifies the link between a class or a category and its members or species*

This relationship must have the mathematical property of inheritance. What is true of a given class is also true of all the classes subsumed under it. The relationship is correct if both the genus and species are of the same fundamental category (facet).

The logical test:

A is a type of B <i>Squirrels are a type of Rodent</i> A is always a type of B <i>Rodents are always a type of mammal</i>
--

The all/some test:

Some members of a class X are entities Y, while all entities Y are members of class X <i>Some members of a class (animals) are entities (vertebrates), while all entities (vertebrates) are members of class (animals).</i>
--

- Hierarchical whole-part relationship (“is part of ...”):  
*One concept is inherently included in another, regardless of context, so the terms can be organized into logical hierarchies, with the whole treated as a broader term*

The relationship is valid *if* the name of the part implies the name of the possessing whole in any context.

Four types have been identified, as follows:

- Systems and organs of the body
- Geographical location
- Discipline or field of study
- Hierarchical social structure

The logical test:

A is an element, subset, aspect, or object of B or vice versa and at least one of the following is true:

If A exists, then B exists.

*If Ottawa exists, then Ontario exists.*

If B exists, then A exists.

*If Ontario exists, then Canada exists.*

In all other circumstances, the whole-part relationship will be an associative one, i.e. the terms will be related terms (RTs).

- Instance relationship (“is an example of...”):  
*Identifies the link between a general category of things or events, expressed by a common noun, and an individual instance of that category*

Example:

*The Rocky Mountains are an example of a mountain range.*

- Polyhierarchical relationship (“is found in ... context”):  
*In which some concepts belong, on logical grounds, to more than one category*

A generic polyhierarchical relationship:

*Organ belongs to both wind instrument and keyboard instrument hierarchies*

A whole-part polyhierarchical relationship:

*Biochemistry is part of biology and chemistry*

Multiple polyhierarchical relationship:

*Skull belongs to bone (kind of), and is also belongs to head (part-of)*

## Chapter 3: Assessing Existing Controlled Vocabularies

GC departments and agencies have access to existing GC controlled vocabularies. Many departments and agencies have developed their own controlled vocabularies. If a department or agency wants to adopt an existing controlled vocabulary or assess their existing ones, a method of evaluation should be selected. This evaluation is essential to determining the feasibility of continuing to develop the existing controlled vocabulary or to adopt an existing controlled vocabulary.

### 3.1 Identifying and Examining the Existing Problem

If there are information management needs that are not met by the current controlled vocabulary then the following questions should be asked:

- Are there any business concepts not represented?
- Is there a problem with ambiguity or redundancy, e.g. are synonyms being used as indexing terms?
- Is there a lack of consistency in format or syntax, e.g. are abbreviations, acronyms, and expanded forms being used variously and indiscriminately together?
- Are lengthy descriptive phrases used instead of more concise and explicit terms?
- Are there problems with search results e.g. lack of precision and recall; too many false drops?
- Are proper names consistently included or excluded?
- Is there overall central editorial control?
- Have users been identified?
- Are users being trained?
- Is the community using the controlled vocabulary as intended?

Once the above assessment is completed and outstanding issues are resolved the organization may move forward to either develop its new controlled vocabulary or to seek an external solution.

### 3.2 Scoping the Solution

The organization must complete the following actions before making a decision whether to improve the existing controlled vocabulary or to build a new controlled vocabulary.

- Establish the scope of the required controlled vocabulary. Identify the principal terminology domains of the organization, including its strategic vision and business focus.
- Determine the intended users.
- Determine the size of the information holdings to be indexed.
- Determine the scope of the information domain, i.e. broad and general or narrow and specific.
- Determine the rate of growth of the information holdings.
- Determine the rate of change or growth of language used in the information domain (e.g. the information domain may be subject to rapid expansion as the result of current research).
- Determine the number of queries directed at information holdings and the specificity of those queries.

In general, the greater the magnitude of each of the above factors, the larger and more complex the vocabulary likely needs to be in order to meet the business requirements.

### **3.3 Selecting Available Controlled Vocabularies**

If you are searching for an already existing controlled vocabulary to meet your organizational needs, you may wish to consider the registered controlled vocabularies found in the LAC and Dublin Core Metadata Initiative web sites, there are also, other controlled vocabularies solutions available on-line and in print.

Initial considerations should include the following:

- Can the vocabulary be extended?
- Is the vocabulary available in both official languages?
- Is it publicly and freely available or is some form of licensing required?
- Are there any other restrictions on its use?
- Has it been well managed and is it likely to be maintained, as both language and terminological needs will evolve over time?
- Is it well documented and does it contain complete usage guidelines?
- Is its terminology generally appropriate in the GC and local departmental or agency contexts?



- Does the controlled vocabulary meet the organization's business requirements?
- Will vocabulary owners consider requests for the addition of new terms or the modification of existing ones?

To address the last point, the owner of a partially useful existing controlled vocabulary should be contacted with a view to exploring the possibility of making changes. Some negotiation may be required, but the final enhanced vocabulary could be more useful and could have greater potential for reuse in the GC as a whole.

### **3.4 Evaluating a Controlled Vocabulary for Use**

Once located, a controlled vocabulary should undergo a rigorous evaluation before it is selected for use. Those under consideration may be evaluated using the following methodology:

- 1) Select a statistically representative set of information resources held by the organization.
- 2) Consider and ensure the appropriate representation of:
  - The size of the information holding
  - The range of resource type
  - The range of subjects and disciplines with which the organization is concerned
  - The group within the organization that creates or uses information resources
- 3) Index a representative set of documents with the controlled vocabulary being evaluated, using a predetermined indexing procedure.
- 4) Assess the results using the following criteria:
  - Are there significant concepts the controlled vocabulary fails to represent?
  - Does the controlled vocabulary use language familiar to the user community of the organization?
  - Is the vocabulary capable of indexing to the required level of specificity?
  - Does the vocabulary include a sufficient number of lead-in terms (i.e. synonyms) that will guide the user community to the appropriate indexing terms, particularly if the vocabulary is large?

- Taking into consideration all factors previously considered or assessed, will this controlled vocabulary serve the information management requirements of the organization?

It may not be possible to modify an existing controlled vocabulary due to the intellectual property rights of the parent organization, costs, or technical considerations. However, it may be possible to acquire the complete vocabulary from the parent organization and modify it independently.

The adoption of an existing controlled vocabulary entails many of the decisions required in the maintenance of a vocabulary produced from the ground up, the guidelines presented here should be reviewed and followed, should this approach be employed. For assistance in carrying out analyses, seek advice from your departmental or agency information management and content experts in your organization.

## **Chapter 4: Design and Construction of New Controlled Vocabularies**

This guide is specific to the development and maintenance of controlled vocabularies in the Government of Canada. It is not intended for the creation, design and construction of controlled vocabularies. Nevertheless, some basic recommendations for the creation of a new controlled vocabulary can be found below.

The methodology used in Chapter 3 for assessing existing controlled vocabularies can be leveraged when the design and development of a new controlled vocabulary is under consideration.

When no pre-existing controlled vocabulary can be adopted or adapted to meet ongoing business and information management requirements, then the creation of a new controlled vocabulary may be necessary. The creation and maintenance of a controlled vocabulary, especially one containing a large number of terms, is a complex and resource-intensive undertaking requiring knowledgeable and experienced personnel and should not be undertaken lightly. Suggested materials for guidance and standards on controlled vocabularies are listed in appendix C.

### **4.1 Scope and Use of the Information Domain**

It is essential to identify the resources to which the controlled vocabulary will be applied, as it is from this information domain that the terms (preferred and non-preferred) will be selected. Determining how you will use your controlled vocabulary will also determine its form.

Review the information domain and define the following:

- Subject area of the controlled vocabulary
- Scope of the subject area
- Level of specificity
- Whether controlled language or natural language or a hybrid will be used
- Purpose driven or content driven

## 4.2 Determine Function, Size and Structure

A key consideration is the function of a controlled vocabulary. Once the function is determined the size and structure can be identified. A vocabulary with a narrowly defined subject area and an extremely small number of terms (e.g. 10–30) that requires no synonyms and is less likely to change over time may be relatively straightforward to develop and maintain. However, if the body of knowledge to be addressed is large, varied, or expansive, then the controlled vocabulary to describe it is more likely to be hierarchical and polyhierarchical.

The basic question is how specific the vocabulary should be to address the identified information needs. A vocabulary that is too general will retrieve too much irrelevant information; one that is too specific will retrieve too little.

The vocabulary should be sufficiently specific to address the information needs identified. “Granularity” is the term used to describe the level of detail reflected by the terms of a controlled vocabulary. A vocabulary of low granularity will represent many concepts under one term; one with high granularity will break out each concept into its own separate terms.

Generally, the broader the scope of the content to be described, the lower the granularity of the vocabulary. A highly focussed subject area usually requires a high degree of granularity. Every controlled vocabulary has a defined range of coverage. When all the possible concepts are included in the content under consideration the controlled vocabulary is then considered exhaustive. Business needs of the domain will always play a part in the scope of the content described.

## 4.3 Concept Representation

It is useful to categorize indexing terms into generalized categories. Below we have identified three main categories and their subdivisions. For a complete specification refer to the ISO 25964-1:2011 ISO standard.

### Concrete Entities

- Things
- Materials

### Abstract Concepts

- Actions and events
- Abstract entities
- Properties of things, materials, and actions
- Disciplines and sciences
- Units of measurement

### **Proper Nouns**

- Individual entities
- Classes of one

# Chapter 5: Management and Maintenance of Controlled Vocabularies

## 5.1 Governance Structure

A fundamental requirement for the use of controlled vocabularies is that ownership and accountability for maintenance and sustainability is clearly documented. This will include instructions on access and procedures for using external vocabularies or explicit policy on the development and maintenance of internal vocabularies supported by clear governance structures.

All vocabulary policies and procedures should be thoroughly documented and continuously updated. Major changes in vocabulary policies or procedures that will affect any users should be communicated to them as soon as possible.

The governance structure for the controlled vocabulary should be established at the outset. This structure includes the following:

- Ownership: roles and responsibility
- Editorial policy
- Indexing policy
- Maintenance policy
- Publication policy

## 5.2 Editorial Policy

Policies should be developed for the controlled vocabulary. Having an editorial policy will manage the direction and evolution of the controlled vocabulary. By managing such issues as duplications, orphans, consistency rules, etc, will provide significant data quality benefits by reducing variations. A controlled vocabulary must change, gradually over time if it is to remain relevant.

The editorial policy should establish the rules for the following:

- Grammatical forms of words or phrases that are used as preferred and non-preferred terms
- Singular or plural forms (depending on the language of the term)
- Punctuation and capitalization

- Abbreviations and acronyms
- Loan words
- Neologisms, slang terms, and jargon
- Common names versus trade names
- Trade marks
- Popular names and scientific names
- Place names (the GC has registered geospatial vocabularies)
- Proper names of institutions, persons, etc. (identifiers)
- Use of warrants to select terms
- Compound terms
- Lexical variants

Example:

Rule Name	Editorial Rule
Amperands	Character & is preferred to the word “and” in term labels. <i>Use cats &amp; dogs Not cats and dogs</i>

### 5.3 Indexing Policy

An indexing policy provides added value through creating additional access points, as well as, providing easier and more precise access to the value domain. Poor indexing practices will degrade the consistency and value of the information being retrieved. This may have serious cost implications.

An indexing policy should be established to address such aspects as:

- Functionality of the content storage, organization and retrieval
- How the vocabulary will be applied to information resources
- Indexing language
- Indexing software that supports the application
- Level of specificity and exhaustivity
- Maximizing the findability of the content
- Recommendations for user friendly design

### 5.4 Maintenance Policy

Maintenance is very important, because if not properly maintained it can affect the reliability of the information retrieved and lead to inconsistencies. Even the simplest controlled vocabulary will need to be maintained as the information domain changes or

expands. Consequent changes in terminology can be labour-intensive and time-consuming. Therefore they must be identified as part of ongoing resource requirements. If the vocabulary is publicly accessible, it is open to public criticism, and liability issues could arise if it is not properly maintained. Maintenance processes need to be specified so that the changes are based on rational cost/benefit decisions, with an awareness of their impact.

An official maintenance policy should be established to address such aspects as:

- How the vocabulary will be managed over time
- How to incorporate user feedback and user search queries into updates
- How new concepts that need to be captured will be incorporated
- How changes to the information requirements will be monitored—these may include changes such as new legislation, modifications to the objectives of the department or agency, or new users and/or purposes for the vocabulary
- How changes within the information domain will be assessed over time
- How terms are added, modified or deleted (specific processes must be identified)
- How changes to terms are tracked over time
- How changes to local technology will affect operation
- Whom to inform when changes are made (the vocabulary may be used by many different users and organizations that need to be informed of modifications)

## **5.5 Publication Policy**

A publication policy will ensure that the availability of the controlled vocabulary is disseminated to the appropriate parties in a suitable manner. It will establish the sharing, accessing, and the re-using of data and any methods for archiving and preserving data.

A publication policy should consider including the following elements:

- Statement of purpose
- Statement of scope
- Identification of ongoing editorial authority and contact information
- Number of indexing terms
- Compliance with standards
- Structure and interrelationships
- Thesaurus [controlled vocabulary] layout and display



- Abbreviations and punctuation
- Operational use
- Updating and maintenance
  - Date last updated

## 5.6 Technology

Technology can facilitate the development of concept knowledge within controlled vocabularies and facilitate the update, maintenance, indexing, interoperability and retrieval within a multilingual environment. It must be compatible with departmental standards and the resources that are being supported.

The maintenance of a controlled vocabulary can be efficiently done when using appropriate software. Auto classification tools can add structure to the data source that can enhance retrieval functionality and information management.

For the most basic controlled vocabularies, such as flat files with few if any synonyms or relationships between terms, common spreadsheet or word processing packages may suffice. However, for more complex requirements, specialized software should be investigated. Some commercial packages are available. It is not the intention of this guide to identify and evaluate specific software, but instead to educate on the subject of controlled vocabularies so that if or when evaluating potential applications readers will be cognizant of needs and be able to be a critical judge when selecting a tool.

In reviewing potential tools, considerations include the following:

- Ensuring the tools will co-exist with the existing and emerging technological environment within the department or agency.
- If the tool is registered and publicly available, the tool must support modification.
- Tools with English and French interfaces and the ability to map linguistic equivalence are preferred.
- If possible, tools compliant with ISO standards should be selected.

The Dublin Core Metadata Initiative maintains a Web site devoted to tools and software (<http://dublincore.org/tools>). See also the Library and Archives Canada's website (<http://www.collectionscanada.gc.ca/government/controlled-vocabularies/007004-7000-e.html>) for information and tools.



## Appendix A: Glossary

Associative relationship ( <i>Relation associative</i> )	Relationship between a pair of concepts that are not related hierarchically but share a strong semantic connection. See “Related term.” (Ref: ISO/DIS 25964-1)
Authority control ( <i>Contrôle d'autorité</i> )	Creating and maintaining terms in a structured manner, so that similar or identical terms can be disambiguated, it is also used to collocate materials that logically belong together, although they present themselves differently.
Authority file ( <i>Fichier d'autorité</i> )	A list or file that is maintained to ensure the consistency of indexing and that establishes the authoritative form of a corporate, geographic or personal name that is to be used to index records. An authority file may contain variant forms of names that are cross referenced to the authoritative form of a name.
Broader term ( <i>Terme générique</i> )	Preferred term representing a concept that is broader than the one in question. (Ref: ISO/DIS 25964-1)
Classification scheme ( <i>Schéma de classification</i> )	Schedule of concepts and pre-coordinated combinations of concepts, arranged by classification. (Ref: ISO/DIS 25964-1)
Compound equivalence ( <i>Équivalence composée</i> )	Relationship between terms or mapping between concepts in which one term or concept in one context is represented by two or more terms or concepts in another. (Ref: ISO/DIS 25964-1)
Compound term ( <i>Terme composé</i> )	Term that can be split morphologically into separate components. Note: Compound terms can be multi-word terms, or can consist of only one word. E.g. “lawnmowers” can be split into “lawn” and “mowers”. (Ref: ISO/DIS 25964-1)
Content driven ( <i>Axé sur le contenu</i> )	Is where the organizational content is used to build a controlled vocabulary which will result in a content driven, bottom-up approach. Usually used for adding terms to an existing vocabulary. Start with terms having the narrowest scope and move up and out towards generic terms.

Controlled vocabulary ( <i>Vocabulaire contrôlé</i> )	Prescribed list of terms, headings or codes, each representing a concept. (Ref: ISO/DIS 25964-1)
Equivalence relationship ( <i>Relation d'équivalence</i> )	Relationship between two terms in a thesaurus that both represent the same concept. (Ref: ISO/DIS 25964-1)
Exhaustivity ( <i>Exhaustivité</i> )	This designates the range of concept coverage of terms in a controlled vocabulary. If the terms cover all of the concepts included in the information domain, then the controlled vocabulary is exhaustive.
Facet ( <i>Facette</i> )	Grouping of concepts of the same inherent category. (Ref: ISO/DIS 25964-1)
Flat file ( <i>Fichier plat</i> )	Simple lists of terminology without synonyms or non-preferred terms. It is the simplest form of a controlled vocabulary.
Folksonomy ( <i>Folksonomie</i> )	A system of classification in which there is no standardization or common vocabulary employed. Individuals collaboratively create and manage tags to annotate and categorize digital content. It is also referred to as collaborative tagging, social classification, social indexing, and social tagging.
Functional vocabulary ( <i>Vocabulaire fonctionnel</i> )	A controlled vocabulary that describes the functions and operations of an organization.
Granularity ( <i>Granularité</i> )	The level of specificity with which content is described by terms in a controlled vocabulary. A controlled vocabulary of low granularity represents many concepts under one term; one with higher granularity breaks out the concepts into their own separate terms.
Hierarchical relationship ( <i>Relation hiérarchique</i> )	Relationship between a pair of concepts of which one has a scope falling completely within the scope of the other. (Ref: ISO/DIS 25964-1)
Indexing ( <i>Indexation</i> )	Intellectual analysis of the subject matter of a document to identify the concepts represented in it, and allocation of the corresponding index terms to allow the information to be retrieved. (Ref: ISO/DIS 25964-1)

<p>Information domain (<i>Domaine d'information</i>)</p>	<p>A well-defined area of knowledge, including the information resources pertaining thereto.</p>
<p>Lead-in term (<i>Non-descripteur</i>)</p>	<p>See “Non-preferred terms.”</p>
<p>Lexical variants (<i>Variantes lexicales</i>)</p>	<p>Words can be classified as lexical or nonlexical. Lexical words are those that have independent meaning (such as a noun, verb, adjective, adverb, or preposition). Lexical variants differ from synonyms in that synonyms are different terms for the same concept, while lexical variants are different word forms for the same expression. These forms may derive from spelling or grammatical variation or from abbreviated formats.</p>
<p>Linked data (<i>Données liées</i>)</p>	<p>Is a method of publishing structured data so that it can be interlinked and become more useful. It builds upon standard Web technologies such as <u>HTTP</u> and <u>URIs</u>, but rather than using them to serve web pages for human readers, it extends them to share information in a way that can be read automatically by computers. This enables data from different sources to be connected and queried. (Ref: Heath, T., Hepp, M., and Bizer, C. (eds.). <i>Special Issue on Linked Data, International Journal on Semantic Web and Information Systems(IJSWIS)</i>).</p>
<p>Literary warrant (<i>Justification par la littérature</i>)</p>	<p>Justification for the inclusion of a term based on the significant frequency of occurrence of those terms in the information resources to be indexed. This assures that resources that could be indexed using the terms in the vocabulary actually exist and terms that do not exist are excluded. This ensures that the form of term selected as the preferred term will be the most common term used in the information resources.</p>
<p>Loan word (<i>Emprunt linguistique</i>)</p>	<p>Term borrowed from another language that has become accepted in the borrowing language. (Ref: ISO/DIS 25964-1) E.g. odour (French), blitz (German), siesta (Spanish).</p>
<p>Mapping (<i>Mise en correspondance</i>)</p>	<p>The process of relating the terms in one controlled vocabulary to those in another.</p>

<p>Metadata (<i>Métadonnées</i>)</p>	<p>The definition and description of the structure and meaning of information resources, and the context and systems in which they exist. (Ref: <i>Standard on Metadata</i> <a href="http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909">http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909</a>).</p>
<p>Narrower term (<i>Terme spécifique</i>)</p>	<p>Preferred term representing a concept which is narrower than the one in question. (Ref: ISO/DIS 25964-1).</p>
<p>Neologism (<i>Néologisme</i>)</p>	<p>A new word coined from an existing term or terms (<i>example</i>: netiquette from "network" and "etiquette"), or a new meaning given to an existing word (<i>example</i>: "quark" from <i>Finnegans Wake</i> used in physics as the name of a subatomic particle). Acronyms are neologisms. Also refers to the use of such a new word or meaning. (Ref: ODLIS <a href="http://www.abc-clio.com/ODLIS/odlis_n.aspx">http://www.abc-clio.com/ODLIS/odlis_n.aspx</a>).</p>
<p>Non-preferred term (<i>Terme non préféré</i>)</p>	<p>Term that is not assigned to documents but is provided as an entry point in a thesaurus or index. (Ref: ISO/DIS 25964-1).</p>
<p>Ontology (<i>Ontologie</i>)</p>	<p>A hierarchical structure that formally defines the semantic relationship of a set of concepts. Used to create structured / controlled vocabularies for the discovery or exchange of information. A thesaurus, like the AAT is an example. (Ref: DCMI <a href="http://dublincore.org/documents/usageguide/glossary.shtml#O">http://dublincore.org/documents/usageguide/glossary.shtml#O</a>).</p>
<p>Organizational warrant (<i>Justification par l'organisation</i>)</p>	<p>Justification for the inclusion of a term in a controlled vocabulary can be based on the specialized language requirements or jargon of the group or organization that is creating or sponsoring a vocabulary.</p>
<p>OWL (OWL)</p>	<p>Web ontology language (<a href="http://www.w3.org/TR/owl-features/">http://www.w3.org/TR/owl-features/</a>). OWL is a language for describing ontologies and schema. It can specify concepts and their relationships. OWL/XDD (XML declaration description) allows a means to express complex rules and constraints. (Ref:DCMI <a href="http://dublincore.org/documents/usageguide/glossary.shtml#O">http://dublincore.org/documents/usageguide/glossary.shtml#O</a>).</p>
<p>Permuted Index (<i>Index permuté</i>)</p>	<p>Also known as Keyword in Context KWIC. Each phrase is indexed by every word used in the phrase. This allows for all cyclic permutations of the phrase to be searched.</p>

Polyhierarchical Structure ( <i>Structure polyhiérarchique</i> )	Hierarchical arrangement of concepts in a thesaurus or classification scheme, in which each concept can have more than one broader concept. (Ref: ISO/DIS 25964-1).
Post-coordination ( <i>Postcoordination</i> )	Combination of preferred terms of a controlled vocabulary at the time of searching. (Ref: ISO/DIS 25964-1).
Pre-coordination ( <i>Précoordination</i> )	Combination of concepts, classes or terms of a controlled vocabulary at the time of its construction or at the time of using it for indexing or classification. (Ref: ISO/DIS 25964-1).
Precision ( <i>Précision</i> )	A measure of a search system's ability to retrieve only relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the total number of content objects retrieved. A high-precision search ensures that, for the most part, the content objects retrieved will be relevant. However, a high-precision search may not retrieve all relevant content objects. <i>See also</i> recall. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down. (Ref: NISO Z39.19-2005).
Preferred term ( <i>Terme préféré</i> )	Term used to represent a concept when indexing. (Ref: ISO/DIS 25964-1).
Purpose driven ( <i>Axé sur les objectifs</i> )	Is where the organizational needs will be used as a methodology, which will result in a top down approach that will be purpose driven. Usually used for new vocabularies.
Quasi-synonym ( <i>Quasi-synonyme</i> )	One of two or more terms whose meanings are generally regarded as different in ordinary usage but which may be treated as labels for the same concept, in a given controlled vocabulary. E.g. diseases, disorders. (Ref: ISO/DIS 25964-1).

Qualifier ( <i>Qualificatif</i> )	A word or phrase used to distinguish a term from otherwise identical terms that have different meanings. A qualifier is separated from the terms, generally displayed within parentheses. It is also called a gloss, although strictly speaking a qualifier should be used only with homographs.
Recall ( <i>Rappel</i> )	A measure of a search system's ability to retrieve all relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the number of all relevant content objects in a collection. A high recall search retrieves a comprehensive set of relevant content objects from the collection. However, high recall increases the possibility that less relevant content objects will also be retrieved. See <i>also</i> precision. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down. (Ref: NISO Z39.19-2005).
Related term ( <i>Terme associé</i> )	Preferred term representing a concept which has an associative relationship with the one in question. (Ref: ISO/DIS 25964-1).
Resource Description Framework (RDF) ( <i>Cadre de description des ressources (RDF)</i> )	The basic language for writing metadata; a foundation which provides a robust flexible architecture for processing metadata on the Internet. RDF will retain the capability to exchange metadata between application communities, while allowing each community to define and use the metadata that best serves their needs. For more information see <a href="http://www.w3.org/RDF/">http://www.w3.org/RDF/</a> (Ref: DCMI <a href="http://dublincore.org/documents/usageguide/glossary.shtml#O">http://dublincore.org/documents/usageguide/glossary.shtml#O</a> ).
Scope note ( <i>Note d'application</i> )	Note that defines or clarifies the semantic boundaries of a concept as it is used in the structured vocabulary. (Ref: ISO/DIS 25964-1).



Semantic factoring ( <i>Factorisation sémantique</i> )	A term which expresses a complex notion is re-expressed in the form of simpler or definitional elements, each of which can also occur in other combinations to represent a range of different concepts. (Ref: ISO/DIS 2788)
Semantic Web ( <i>Web sémantique</i> )	Provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. (Ref: W3C <a href="http://www.w3.org/RDF/FAQ">http://www.w3.org/RDF/FAQ</a> ).
Specificity ( <i>Spécificité</i> )	Capability of a structured vocabulary to express a subject in depth and in detail. (Ref: ISO/DIS 25964-1).
Synonym ( <i>Synonyme</i> )	One of two or more terms denoting the same concept. E.g. guarantees, warranties (Ref: ISO/DIS 25964-1).
Syntactical factoring ( <i>Factorisation syntaxique</i> )	Applied to compound terms which are amenable to morphological analysis into separate components, each of which can be accepted as an indexing term in its own right. (Ref: ISO/DIS 2788).
Syntax ( <i>Syntaxe</i> )	The form in which the terms of a controlled vocabulary are structured and represented. For example, <i>last name first, comma, first name, comma, middle initial, period</i> .
Taxonomy ( <i>Taxonomie</i> )	A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy. (Ref: ISO/DIS 25964-1).
Term ( <i>Terme</i> )	A word or phrase used to label a concept. (Ref: ISO/DIS 25964-1).
Thesaurus ( <i>Thésaurus</i> )	Controlled and structured vocabulary in which concepts are represented by terms, organized so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms. (Ref: ISO/DIS 25964-1).

Top term  
(*Terme supérieur*)

Preferred term representing a concept that has no broader concept in the thesaurus. (Ref: ISO/DIS 25964-1).

User warrant  
(*Justification par l'utilisation*)

This term refers to a justification for selecting terms based on words or phrases employed by users of information resources for information retrieval or information management. Evidence of such usage may be derived from search engine logs.

Vocabulary control  
(*Contrôle de vocabulaire*)

Management of a vocabulary in order to disambiguate and constrain the form of the terms and limit the number of concepts and terms available for indexing. (Ref: ISO/DIS 25964-1).

## Appendix B: Selected Bibliography

Aitchison, J.A.; Gilchrist, A.; Bawden, D. *Thesaurus construction and use: A practical manual*. 4th ed. Chicago, IL., Fitzroy Dearborn, 2000. Print.

Brown, Fred. *Vocabulary Links: //Thesaurus Design for Information Systems – Seminar by Dr. Bella Hass Weinberg*. 1998. Web. 6 November 2012 <<http://www.allegrotechindexing.com/article02.htm>>.

Chan, Lois Mai; Zeng, Marcia Lei. *Ensuring Interoperability among Subject Vocabularies and Knowledge Organization Schemes: A Methodological Analysis*. 68th IFLA Council and General Conference, August 18–24, 2002 Proceedings. Web. 6 November 2012 <<http://www.ifla.org/IV/ifla68/papers/008-122e.pdf>>.

Craven, T. *Thesaurus Construction*. London, Ont., University of Western Ontario. 2001. Web. 6 November 2012 <<http://publish.uwo.ca/~craven/677/thesaur/main00.htm>>.

Getty Research Institute. *Art and Architecture Thesaurus Online*. 2000. Web. 15 November 2012 <<http://www.getty.edu/research/tools/vocabularies/aat/>>.

Hedden, Heather. *The accidental taxonomist*. Medford, N.J.: Information Today, Inc, 2010. Print.

Hudon, Michèle. “Relationships in multilingual thesauri.” *Relationships in the Organization of Knowledge*. (2001): 67–80. Print.

International Organization for Standardization (ISO). ISO 25964-1:2011 *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. Geneva, Switzerland: International Organization for Standardization.

Lambe, Patrick. *Organising knowledge: taxonomies, knowledge and organisational effectiveness*. Oxford, U.K.: Chandos Publishing, 2007. Print.

Harper, Corey A. and Barbara B. Tillett. “Library of Congress Controlled Vocabularies and their application to the semantic Web.” *Cataloging & Classification Quarterly* 43.3-4 (2007): 47-68. Print.

National Archives of Australia. *Developing a Functions Thesaurus*. 2003. Web. 6 November 2012 <[http://www.naa.gov.au/Images/developing-a-thesaurus\\_tcm16-47228.pdf](http://www.naa.gov.au/Images/developing-a-thesaurus_tcm16-47228.pdf)>.

National Information Standards Organization (NISO). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. [ANSI/NISO Z39.19-2005]

## Appendix C: Guidance and Standards

Aitchison, J.A.; Gilchrist, A.; Bawden, D. *Thesaurus construction and use: A practical manual*. 4th ed. Chicago, IL., Fitzroy Dearborn, 2000. Print.

International Organization for Standardization (ISO). ISO 704-3:2009 *Terminology Work – Principles and Methods*. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 860:2007 *Terminology Work – Harmonization of Concepts and terms*. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 1087-1:2000 *Terminology Work – Vocabulary – Part 1 Theory and Application*. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 13250:2003 *Information technology – SGML applications – Topic maps*. Geneva, Switzerland: International Organization for Standardization.

International Organization for Standardization (ISO). ISO 25964-1:2011 *Information and Documentation – Thesauri and Interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. Geneva, Switzerland: International Organization for Standardization.

National Information Standards Organization (NISO). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. [ANSI/NISO Z39.19-2005]

National Information Standards Organization (NISO). *Guidelines for Indexes and Related Information Retrieval Devices*. [NISO TR-02-1997]

Treasury Board of Canada Secretariat. *Standard on Metadata*. July 2010. Web. 6 November 2012 < <http://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=18909>>.