Governments meet people: Developing metathesauri in the framework of “government online” initiatives

Abstract: An objective of government online initiatives is to offer publicly available government information and services on the Internet. A potential solution to the difficulties experienced by citizens-customers in their search for pertinent information and services is the development and application of a metathesaurus. The metathesaurus is a high-level, interdisciplinary, structured list of terms representing concepts relating to the government administration itself, and including also broad descriptors of a government’s main areas of responsibilities. Five such vocabulary products are described in this paper.

Résumé: Les projets dits de gouvernement en ligne ont comme objectif d’offrir sur Internet de l’information et des services gouvernementaux de nature et d’intérêt publics. Une solution possible aux problèmes rencontrés par les citoyens dans leur recherche d’information pertinente est le développement et l’application de métathésauri. Le métathésaurus est une liste structurée et interdisciplinaire de termes généraux qui décrivent des concepts directement reliés à l’administration gouvernementale elle-même et qui inclut également des descripteurs provenant de domaines au sein desquels un gouvernement intervient de façon significative. Cinq vocabulaires de ce type sont décrits dans cet article.

1. E-GOVERNMENT AND GOVERNMENT ONLINE INITIATIVES

The trend for governments around the world to load huge quantities of information on web sites for all citizens to see and use is the most visible manifestation of a phenomenon labelled “e-government” or “government online”. Electronic government refers to a “government’s use of technology, particularly web-based Internet applications, to enhance the access to and delivery of government information and service to citizens, business partners, employees, other agencies, and government entities” (Layne and Lee, 2001, 123, quoting McClure, 2000). The objective is for information and resources to “flow seamlessly within and between all public sector bodies” (Dextre Clarke, 2002), hopefully becoming in the process more accessible to citizens. The governments of the United Kingdom and Canada aim at full interactivity by the year 2005.

Allen and others suggest that digital government “be viewed as much more than moving existing public services on-line: it is about government harnessing IT to redefine its
"social technologies" in order to remain relevant in a more participative, more interactive, and more informational era" (Allen et al., 2001, 94). E-government is related to the goal of e-governance, a concept of interest to political scientists. E-governance calls for a complete restructuring of government and a profound transformation of its internal culture. The discussion as to whether or not such a radical change is realistic obviously falls outside the limits of this paper although we will occasionally refer to some of its main arguments.

A four-stage model for implementation of e-government is proposed by Layne and Lee (2001). During the first stage, the cataloguing stage, various agencies within a government establish an online presence, usually by loading on the Web information deemed of interest to their own employees and to citizens. Government online initiatives are set-up at this stage. Two-way communication is established at the transaction stage. The focus has now moved to designing on-line interfaces that allow citizens to transact with government electronically if they so desire. Access is through a portal that reflects customers needs rather than the complex structure of the administration. Vertical integration happens at the third stage, when various levels of government (federal, provincial, and municipal) connect to offer similar or related services (for example, the issuing of licenses and permits). At this stage, the focus has moved beyond automation and digitization of existing processes toward transformation of how governments do business with each other and with taxpayers. The ultimate stage, that of horizontal integration, is defined as integration across different functions and services within the same government. Databases become able to "communicate with each other and ideally share information, so that information obtained by one agency will propagate through out all government functions" (Layne and Lee, 2001, 132). Reaching stage four may or may not be possible for a government: it has not happened yet, as it requires an inside out transformation of government functions to more service oriented ones. In Canada, for example, the pledge of "one-stop shopping and service integration implies a degree of horizontal coordination and information sharing that is presently neither legislatively permissible nor quite likely in a cultural context shaped largely by traditional public service values linked to Ministerial (read vertical) accountability" (Allen et al., 2001, 95).

At this time, if very few government agencies remain without their own web site, fewer can offer anything approaching full electronic service delivery (Paschoud, 2000), and this even if considerable sums of money have already been spent on government online initiatives. It is clear, however, that if the 1990's can be seen as the decade of technology, we have now moved into the decade of content.

2. RESOURCE DESCRIPTION AND VOCABULARY ISSUES

Making the information available is of course not the same as making it accessible. Navigating government web sites can result in both satisfaction and frustration: satisfaction with the rich information available, but frustration in trying to locate a specific document or service (Moen, 2001, 155). Fortunately, some individuals have realized quickly the importance of structured access to the content of government web
sites. The terms metadata, and somewhat surprisingly, taxonomies, thesauri, topic trees, etc. have now invaded the discourse of those responsible for government online initiatives. All agree that "successful interoperation between citizens and a myriad of government information systems [depends] critically on the establishment, wide circulation, and agreement of metadata standards by which all can connect" (Paschoud, 2000, 285).

A widely used standard is the Government Information Locator System (GILS), implemented in the United States in the mid-1990's. The GILS concept focuses on the use of metadata to enable agency-based locator services to assist users in identifying and acquiring government information (Moen, 2001). At GILS core is a set of approximately 70 elements that agencies can use to describe resources in a uniform and coherent fashion; many of these match essential components of the traditional bibliographic description.

Description of content on the basis of subject is provided for in the original GILS and in most of its variations, without being necessarily a mandatory element. It is becoming more and more obvious, however, that in a user-centric context, the possibility of searching for resources by topic is essential, and vocabulary issues then become critical (Miller, 2000). When searching by subject, even with the support of powerful technology and careful interface design, natural language is not always the most efficient tool for resource discovery.

Westcott describes the civil service as "made up of groups of specialists in a wide variety of specialisms each of which has its own jargon and way of using similar terms" (2000, 2). Within a government, "policy makers do not speak the same language as statisticians, who do not speak the same language as sociologists, who do not use terms in the same way a criminologist will, [and] none speak the same language as the general public"(Westcott, 2000, 2). A solution appears to be the imposition of a degree of control over the terms used by all parties when they describe the content of resources made available to external users. At its most basic, vocabulary control will involve little more than establishing a finite list of words from which application and users will have to select. In more complex instances, fully formed thesauri may be employed (Miller, 2000).

Sections 4 and 5 below provide an overview of choices made by various governments in relation to subject metadata, and describe briefly the main characteristics of language instruments used for content description and indirectly for subject searching.

3. THE METATHESAURUS CONCEPT

Miller believes that thesauri are an "important foundation in effectively creating, curating, and re-using rich information resources in an on-line environment" (2000). Owing to the great popularity of this instrument over the past decades, specialized thesauri are in fact already used in many government agencies. The take-up of thesauri at a government-wide level, however, has been cautious, "doubtless due to their perceived
complexity, and to their relative lack of availability online or as part of off-the-shelf tools" (Miller, 2000).

The value and practicality of creating a single high-level thesaurus, usable across a wide range of on-line resources to increase semantic interoperability, has been recognized (Miller, 2000). In the UK, for example, the development of a pan-governmental thesaurus has been seriously considered (see 4.1 in this paper), while Canada has moved ahead with its transformation of a departmental thesaurus into a standard tool for use in a much wider variety of contexts (see 5 below).

The development of a thesaurus for government-wide application, of necessity an interdisciplinary thesaurus, may be the obvious ideal solution, but that does not make it an easy instrument to create. The complexity of the thesaural structure is an issue indeed, but information specialists with a knack for terminology and semantics can work efficiently toward the production of a tool which can then be built into systems so that users can benefit from controlled terminology without having to understand the ins and outs of vocabulary control.

Problems relating to interdisciplinarity and to the heterogeneity of audiences seem more critical. Governments assume responsibilities in all areas of life, whether individual, social, economic, cultural, etc. To describe all of a government’s interest or fields of action, concepts from numerous disciplines must be represented in the thesaurus. At the lexical level, problems of levels of language stem from the wide variety of potential users: terms used by specialists in a department as well as terms used by citizens to name the same concept need to be identified and listed. In terms of technology, government online initiatives are generally geared to low-end users; should the vocabulary used for citizen-government communication also be at a lower level than it has been in the past?

Another challenge is the necessity for a government-wide thesaurus to retain a high-degree of compatibility with other thesauri used within the same administration, or even outside of it.

Although it is evident that no single thesaurus or other terminological tool will ever meet all the needs of all users all the time, we suggest that the use of a metathesaurus to support resource description and discovery in governments’ web sites may present some interest.

The concept of metathesaurus has been present for a while in the literature of our field, but it has been described in quite different ways. In 1990, Roulin imagined a European metathesaurus of education which he described as a collection of sub-thesauri, co-thesauri, and national vocabularies revolving around the Thésaurus européen de l’éducation, and allowing some form of communication between the specialized vocabularies used in European countries to organize and retrieve education-related information. Howarth (1996) described an interdisciplinary metathesaurus which would facilitate searching across disciplines on the Web. Also depicted as a metathesaurus, one of the three knowledge bases combining to form the Unified Medical Language System.
(UMLS), contains the 776,940 concepts and 2.1 million terms found in over 60 different biomedical source vocabularies (National Library of Medicine, 2002). Depending on the context, it then appears that a metathesaurus can be an interdisciplinary thesaurus, a macro-thesaurus, a super-thesaurus, or a high-level thesaurus. The metathesaurus is a tool that exhibits the "look and feel", but not all of the essential characteristics of the ISO-compliant thesaurus. The term metathesaurus could be used at this time, and most likely temporarily, to represent what the thesaurus, a tool in transition, may in fact become.

Our own vision of a metathesaurus is that of a high-level, interdisciplinary, and structured list of terms, applicable in diverse environments / institutions which do not necessarily share the same mandate, range of action, or even topical interests, to describe the content of resources, on the model and a scale similar to those of the EUROVOC Thesaurus [http://europa.eu.int/celex/eurovoc/]. EUROVOC is a multilingual high-level thesaurus covering the fields in which the European Communities are active. EUROVOC organizes concepts and terms in 21 general categories (Finance, Social questions, Transport, Environment, etc.) and 127 micro-thesauri, each of which can be maintained separately.

A metathesaurus for government-wide application must include two categories of terms. The first category includes terms relating to or representing:

1. The Government itself, its structure, functions, programs and services (For example: Agents: Public servants, Senators, Crown Corporations; Processes: Elections, Government spending; Entities: Government information, Government programs, Bilingual government services);
2. Law-related concepts (For example: Legal aid; Court decisions; Criminal justice; Fiscal law);
3. General management concepts (For example: Agents: Employees, Supervisors, Auditors, Not-for-profit organizations; Processes: Accounting, Staffing, Strategic planning, Strikes, Training, Advertising, Purchasing; Entities: Job applications, Budget, Guidelines);
4. Communication and information management concepts (For example: Agents: Gestionnaire de documents, Webmestre; Processes: Accès à l’information, Classification; Entities: Archives, Document officiel, Répertoire, Intranet);
5. Technology related concepts (For example: Internet; Office automation; Technology transfer);
6. Generic / Function terms (For example: Assessment, Comparative analysis, Innovation, Maintenance, Research).

Terms in the second category are topical terms that describe the responsibilities of governments in various areas: Cultural and Societal affairs, Economy, Public health and safety, Social security, Natural resources and Environment, Transport, International relations, Defence.

To produce and use such a tool efficiently, consensus is needed at two levels:

1. At a government-wide level, on a common terminology to represent government agents, functions, products, services;
2. Among agencies working in overlapping areas, on a common terminology to represent common concepts.

4. VOCABULARY PRODUCTS IN USE OR UNDER DEVELOPMENT

Most of the following vocabulary products are in the final stages of development. All products may be seen as metathesauri.

4.1 Australia

Two thesauri will soon be available to civil servants and to the citizens of Australia.

Designed by the National Archives of Australia, the Australian Government’s Interactive Functions Thesaurus (AGIFT) [http://www.naa.gov.au/recordkeeping/gov_online/agft/extract.html] is a web-based thesaurus that links plain English (300,000 natural language words) with government terminology (close to 600 terms). Although its main objective is to provide standard terms for government agencies to use in the function metadata of the Australian Government Locator Service (AGLS), it can indirectly assist those searching government sites when they don’t know which terms to use or more generally which level of government has responsibility for the service or information required.

*AGIFT* assigns descriptors to one of 24 categories representing major functions of government (Business management, Civic infrastructure, Communications, Community services, Defence, etc.) *AGIFT* also includes a useful description of each sector’s responsibilities.

*TAGS, The Thesaurus of Australian government subjects* [http://www.govonline.gov.au/projects/standards/TAGS/TAGS.htm], is being developed in the framework of the government’s portal project. Its content represents the scope of government activities from a topical rather than from an administrative perspective. *TAGS* can function as a high-level point of entry into the specialized vocabularies used by some agencies or serve as controlled vocabulary for those agencies without a specialized one. Terms are grouped in 15 categories, with hierarchies developed to the fourth level. Categories include Agriculture, Business and industry, Defence, Employment, Government and politics, Law and justice, Transport.

It is not clear at this time if and how *AGIFT* and *TAGS* will interact with each other. Presumably, the functions thesaurus has been judged necessary to preserve the integrity of internal records management operations.

4.2 New Zealand

In the framework of its Citizen’s Portal Thesaurus Project, New Zealand has also adopted the dual thesaurus approach, creating a *Functions of New Zealand (FONZ)* Thesaurus

Agencies supplying metadata for the Portal are required to include at least one term from each of the Functions Thesaurus and the Subjects Thesaurus in the description of each resource. The dual thesaurus model allows the indexers to create their own unique combinations of functions and subjects (For example: Policymaking (F) regarding School zoning (S); Programming (F) Vaccinations (S)), the two thesauri acting as a post-coordinate one. A certain amount of conceptual and lexical overlap is to be expected, as functions may also be discussed in resources and become topics of interest to citizens. FONZ is a small but structurally complex instrument which details government functions under 12 broad headings: Administering government, Assisting, Authorizing, Enabling, Enriching, Informing, Policymaking, Protecting, Providing infrastructure, Representing, Rulemaking, and Upholding the rules. Its structure does not immediately bring to mind that of the ISO-compliant thesaurus.

On the other hand, SONZ lists approximately 2 000 terms in a traditional alphabetical thesaurus display; the terms do not appear to have been assigned to broad categories.

4.3 United Kingdom

In the United Kingdom, a pan-governmental thesaurus (PGT) was first proposed as a workable solution to the difficulty of providing controlled subject access to the main government web site. Having thoroughly examined five other options, and among these the option of not controlling subject access, a group of information professionals eventually settled on a less complex product. The Government Category List (GCL) [http://www.govtalk.gov.uk/documents/GCLv1_0.pdf] is "a small and simple taxonomy, designed to facilitate high-level browsing rather than deep searching" (Dextre Clarke, 2002), which may be complemented by specialized thesauri covering specific domains. The GCL is described as a hybrid between a thesaurus and a classification scheme, capable of ensuring a reasonable level of consistency from the hundreds of people who will use it for meta-tagging. The GCL currently lists over 1 000 lead-in terms and close to 400 descriptors arranged in 12 broad and widely encompassing hierarchies / categories (For example: Agriculture, environment and natural resources; Arts, recreation and travel; Crime, law, justice and rights; People, communities, and living).

The GCL differs from the ISO-compliant thesaurus by the width of its coverage and the fact that its hierarchical structure is not necessarily of the genus-species type. Furthermore, the GCL is polyhierarchical, allowing more than one hierarchical access route to the same subject. The choice of terminology and structure do not reflect the priorities of government employees but rather those of citizens-customers.
The *GCL* is designed for high-level description of resources, most of which will also be described more specifically in originating agencies. To avoid the necessity of indexing the same resources using a specialized vocabulary and then the *GCL*, it will be essential for each agency to develop mapping techniques so that *GCL* terms are automatically selected, rather than humanly assigned, on the basis of the more specific terms used to represent the content of resources.

### 4.4 United States

It is within state libraries (Washington, Oregon, Texas, Illinois, etc.) that the movement for vocabulary control in government web sites developed in the United States. The *Jessica Tree*, designed in the framework of the *Find-It! Illinois* project and named for its developer Dr. Jessica Milstead, has become the nucleus of a more comprehensive topic tree now ready for use by the State GILS community, and at the federal level.

The *GILS Topic Tree* [http://www.fidocat.com/GILS/two.pdf] is made of 669 preferred terms and 692 non-preferred terms, arranged in 26 somewhat overlapping categories (For example: Agriculture and food production; Business and industry; Land use, development and construction, or Federal government; Government finance and taxes; Voting and elections). One category relates to audiences (Kid’s page), another serves as a repository for the various forms that are or are likely to become available online (Accident report forms, insurance forms, etc.).

The tree is described as a network, but also as a “a hierarchical thesaurus that provides terms broad enough at the top levels to accommodate uniform application while extensible at the lower, narrower levels so that all states can customize the thesaurus to add locally unique headings” (Moen, 2001, 158).

Not surprisingly, hierarchical relationships in the *Topic Tree* are not ISO 2788-compliant. Although the tree looks and can be used like a standard thesaurus, it more closely resembles a classification scheme where its relational pattern is concerned. Polyhierarchy is seen as essential, and most concepts will be found in various categories and in various places within the five-level deep hierarchical structure.

### 5. THE GOVERNMENT OF CANADA’S **CORE SUBJECT THESAURUS (CST)**

In Canada, the Treasury Board Secretariat leads the Government On-line Metadata Working Group, established in the summer of 2000 and ongoing at this time. An early initiative of the Group has been to develop two standards, one for resource description, and a second one addressing terminological issues.

The Government online metadata standard [http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit391_e.asp] is based on the Dublin Core. Five elements of description are mandatory: Title, Creator, Language, Date, and Subject. Resources to be described include Welcome pages and home pages, pages that provide actual on-line
service to the public, pages required to meet a prescribed legal or service obligation, major formal publications, information about agency powers affecting the public, and documents used in decision-making affecting the public. It is understood that the metadata standard can be extended or mapped to ensure an even higher level of both consistency and interoperability where needed.

The controlled vocabulary standard [http://www.cio-dpi.gc.ca/its-nit/standards/tbits39/crit392_e.asp] adopts the principle of using controlled vocabulary for the management of subject access to electronic information. More specifically, it designates the Government of Canada Core Subject Thesaurus (CST) [http://dsp-psd.communication.gc.ca/Thesaurus/index-e.html], a high-level, multi-disciplinary bilingual tool, as the default thesaurus to be used by federal organizations for the description of resource contents. This Treasury Board decision owes a lot to the existence, since 1992, of a general thesaurus used to describe the content of Canadian government publications distributed to depository libraries, combined with a belief that the use of a well-structured thesaurus as subject authority file remains desirable to facilitate user searches within a web site and comprehensive retrieval of available sources of pertinent information. The thesaurus is considered capable of linking the natural language of citizens to the structured categorization schemes reflecting the internal structure of the government itself, a structure with which the citizen is not normally familiar.

A first list of descriptors was developed between 1990 and 1992 by the Depository Services Program, then part of Canadian Government Publishing. The list was a vocabulary control instrument intended for use in electronic databases and print catalogues and indexes maintained by the Program. The List of subject descriptors included just over 1,100 terms in each one of its French and English versions. Subject coverage was already far-ranging since terms were needed to represent the variety of subject matter discussed in resources created and distributed in various forms by the Canadian government. Both English and French versions of the tool were developed simultaneously, with neither language seen as dominant or source language. The thesaurus lexicon had reached 1,500 descriptors and about the same number of lead-in terms when an official revised version, appearing under its current title, was made available in 2000. The 2002 version of CST contains close to 4,000 terms, half of which are descriptors. All fields of knowledge are represented in the thesaurus, with emphasis on the following areas: Government and politics, Economics and industry, Nature and environment, and Society and culture. These domains are covered in greater depth, and are represented by a larger number of descriptors. Names of individuals, places, organizations and events are excluded from CST.

CST was developed and is maintained in accordance with the Guidelines for the establishment and development of monolingual thesauri (ISO 2788-1986), and the Guidelines for the establishment and development of multilingual thesauri (ISO 5964-1985). It is thus fully compatible with existing thesauri in specialized fields. CST offers an equivalent in the other language for each descriptor, and provides historical notes, application notes, and a few definitions. Its relational structure is made of equivalence,
hierarchical and general associative relationships. Hierarchical relationships are of a
generic nature only, and most hierarchies are four or five-level deep. The English and
French versions are identical as to their coverage and number of descriptors listed. Their
respective relational structures are comparable but not identical, due to conceptual and
lexical differences between English and French.

The thesaurus also exhibits a superstructure made of 19 categories, 16 of which represent
main areas of government involvement (Agriculture, Economics and industry, Education
and Training, Government and politics, Health and safety, Labour, etc.). Two categories,
Persons and Processes, are strictly functional. The former is used to specify target
audiences while the latter is a list of general terms representing actions, properties, etc.
Examples from the Processes category are: Abilities, Advisory committees, Comparative
analysis, Efficiency, Maintenance, Mandate, etc. The last category lists terms that are
more directly related to the document as physical object, describing either its external
form (Multimedia, Braille, etc.) or its type (Aeronautical charts, Bibliographies, etc.)
Terms are frequently assigned to more than one category: Air transport industry for
example, belongs to both Economics and industry and Transport categories.

Within the federal government, users of CST are primarily those tasked with building,
enhancing or revising the Government of Canada’s various web sites. They are web
developers, librarians, content providers, indexers, web technologists and
communications personnel.

A Thesaurus Working Group Sub-Group on Terminology for Departmental Business
Line has been established to determine the extent to which CST meets the needs of
government departments and agencies for terminology describing what they do, and to
identify candidate terms for inclusion in the main thesaurus. Sixteen departments
submitted lists of high-level terms describing their core responsibilities. 63 % of these
terms were found to be present in CST as either preferred or non-preferred terms. At this
time, a second round of consultations has been initiated, and further terms are being
evaluated for inclusion.

Six departments so far have opted to use CST as their main indexing tool. Other agencies
continue to use locally developed vocabularies. Some of these tools are extensive and
remarkably well-structured. Well before the set-up of a government online initiative,
Statistics Canada, for example, had developed its own ISO-compliant bilingual thesaurus
to index its full suite of publications and surveys. The web-based thesaurus
[http://www.statcan.ca:80/english/search/thesaurus.htm] is fully compatible with CST.

The TBS controlled vocabulary standard specifies that only terms from an authoritative,
registered source can be used in the dc.subject field. The National Library of Canada has
undertaken to establish a Registry based on a model developed by the Dublin Core
initiative. The rationale for the Registry is two-fold: to identify controlled vocabularies in
use within the Government of Canada with a view to making them available to search
engines, web sites users, information creators and scheme developers and maintainers,
and to provide a centralized mechanism for identifying names for use as the dc.subject scheme qualifier.

As CST expands, decisions as to which concepts and terms should be included / excluded are becoming increasingly difficult to make. Specific departments which have not developed controlled vocabularies for their own purposes expect from CST a level of specificity that is not appropriate in a high-level metathesaurus. Where the names of species of animals or vegetable have been included for example, it is to provide a model for further development of the thesaurus, but it does increase the risk that too many specific terms will be added and that great differences in coverage between one area and another will appear. Furthermore, a high level of specificity does not necessarily correspond to citizens-customers needs.

Although CST has been chosen as core thesaurus, it is not visible on the Government of Canada main web site. The thesaurus is available on the Depository Service Program site and it is not yet being marketed directly to citizens as a tool that may facilitate access to government information.

6. DISCUSSION

All controlled vocabularies described above have been developed for use in similar environments: the main portals providing access to governments’ information and services. Although they may never have to communicate with each other, it is still interesting to compare their contents and structure and to evaluate their level of compatibility.

These controlled vocabularies share several characteristics.

- They cover a very wide area of knowledge.
- They try to adopt a level of language closer to that of citizens than that of civil servants.
- They are designed for post-coordinate indexing.
- They are amenable to mapping from specialized vocabularies.
- They exhibit a shallow hierarchical structure.

The controlled vocabularies differ mainly in their degree of compliance with ISO standards for the development of indexing languages. While CST, the Australian thesauri and the Subjects of New Zealand Thesaurus generally conform to ISO 2788, the GCL and the GILS Topic Tree bend the rules and dodge the limitations imposed by the standard’s strict definition of a hierarchy.

All products but one exhibit a superstructure made of a number of broad categories, from a low of 12 to a high of 26. As all governments share the same general types of responsibilities, one expects that these categories will more or less correspond. Although all tools are indeed rather similar in content, Table 1 shows that distinctions, emphasis
and groupings at the highest structural level are quite different. Within its 26 categories, the Topic Tree introduces distinctions between levels of government, for example, that CST and GCL have not considered necessary. On the other hand, CST and the GCL propose a Science and technology class, as could be expected, and it is difficult to see where terms relating to these important concerns of government policy will be found in the Topic Tree, which does not propose a similar category.

<table>
<thead>
<tr>
<th><strong>GILS TOPIC TREE</strong></th>
<th><strong>CST</strong></th>
<th><strong>GCL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture and food production</td>
<td>Agriculture</td>
<td>Agriculture, environment and natural resources</td>
</tr>
<tr>
<td>Business and industry</td>
<td>Economics and industry Labour</td>
<td>Business and industry</td>
</tr>
<tr>
<td>County government</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Education</td>
<td>Education and training</td>
<td>Education, career and employment</td>
</tr>
<tr>
<td>Health and medicine</td>
<td>Health and safety</td>
<td>Health, nutrition and care</td>
</tr>
<tr>
<td>Kids’ pages</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Law enforcement and courts</td>
<td>Law</td>
<td>Crime, law, justice and rights</td>
</tr>
<tr>
<td>Laws and regulations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social issues and programs</td>
<td>Society and culture</td>
<td>People, communities and living</td>
</tr>
<tr>
<td>–</td>
<td>Science and technology</td>
<td>Science, technology, and innovation</td>
</tr>
</tbody>
</table>

Table 1. Selection of categories in the GILS Topic Tree, the Core Subject Thesaurus, and the Government Category List

And what about levels of compatibility beyond that of main classes? The category Education, present in all vocabularies, provides an interesting sample.

Table 2 illustrates significant differences in the lists of top terms offered in the Topic Tree, CST and the GCL within the class Education. The Topic Tree offers the greatest choice of top terms, with wide variations in hierarchical levels and conceptual extension (Athletic associations, Higher education, Libraries and archives, for example).
<table>
<thead>
<tr>
<th><strong>GILS TOPIC TREE</strong></th>
<th><strong>CST</strong></th>
<th><strong>GCL</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult and continuing ed.</td>
<td>Access to education</td>
<td>Adult ed. and training</td>
</tr>
<tr>
<td>Athletic associations</td>
<td>Education</td>
<td>Colleges and universities</td>
</tr>
<tr>
<td>Bilingual education</td>
<td>Educational policy</td>
<td>Pre-school learning</td>
</tr>
<tr>
<td>Directories of ed. institutions</td>
<td>Education institutions</td>
<td>Schools</td>
</tr>
<tr>
<td>Distance learning</td>
<td>Educational resources</td>
<td>Special education needs</td>
</tr>
<tr>
<td>Early childhood education</td>
<td>Educational technology</td>
<td></td>
</tr>
<tr>
<td>Education pages for kids</td>
<td>Learning</td>
<td></td>
</tr>
<tr>
<td>Education programs</td>
<td>Literacy</td>
<td></td>
</tr>
<tr>
<td>Educational finance</td>
<td>Numeracy</td>
<td></td>
</tr>
<tr>
<td>Educational policies</td>
<td>Scholarships</td>
<td></td>
</tr>
<tr>
<td>Educational technology</td>
<td>School boards</td>
<td></td>
</tr>
<tr>
<td>English as a second language</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracurricular activities</td>
<td>Student loans</td>
<td></td>
</tr>
<tr>
<td>Higher education</td>
<td>Students</td>
<td></td>
</tr>
<tr>
<td>Home schooling</td>
<td>Studies abroad</td>
<td></td>
</tr>
<tr>
<td>Libraries and archives</td>
<td>Teachers</td>
<td></td>
</tr>
<tr>
<td>Literacy</td>
<td>Training</td>
<td></td>
</tr>
<tr>
<td>Private schools</td>
<td>Training centres</td>
<td></td>
</tr>
<tr>
<td>Public schools</td>
<td>Vocational guidance</td>
<td></td>
</tr>
<tr>
<td>School personnel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test scores</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Top terms in the category Education / Training**

At first glance, it appears that all vocabularies, beyond the main class structure, exhibit significant differences, no doubt attributable in part to the important degree of subjectivity introduced by vocabulary specialists working without clear standards to guide them in the development of a new type of controlled vocabulary. It is fair to
assume that structures which are not ISO-compliant may exhibit even greater levels of subjectivity than those which are.

7. CONCLUSION

It has now been recognized that technically connecting resources is far easier than actually getting them to work together in any meaningful fashion (Miller, 2000). Whatever product is selected, it will be a challenge to achieve vocabulary control for improved access to government information and services. Dextre Clarke rightly observes that “there is a big contrast between coordinating document input procedures in one organization, and procuring the compliance of all the bodies in the public sector” (Dextre Clarke 2002). Cooperation from all parties involved is key to reaching the objective of basic lexical and semantic interoperability.

None of the instruments described above has yet been fully implemented. At this time, more details are available on their recommended use in connection with resource description operations than with their applicability to the resource discovery process. Citizens have not yet had a chance to use, knowingly or without realizing it, these metathesauri. It will be some time before we can actually evaluate whether metathesauri developed for government-wide application are in fact useful to citizens-customers, and bring them closer to their governments. This remains, let’s not forget, the main objective of government online initiatives.

ENDNOTES

1 “Le Méta-thesaurus est l’ensemble composé du Thésaurus européen de l’éducation et des différents langages d’indexation qui seront rattachés à celui-ci, qu’il s’agisse de sub-thésaurus, de vocabulaires spécialisés, de co-thésaurus ou de vocabulaires nationaux” (Roulin, 1990, 81).

REFERENCES


Core Subject Thesaurus. [http://dsp-psd.communication.gc.ca/Thesaurus/index-e.html]. Accessed 04/14/02.


