

# **Integrated Public Sector Vocabulary**

*Editorial Policy Version 1.2*

*25 October 2004, revised 19 April 2005*

Stella G Dextre Clarke  
Information Consultant

## Document History

This document sets out editorial policy for the Integrated Public Sector Vocabulary. The first edition was approved by the Vocabulary Merger Project Board on 1 November 2004, following discussion of the issues and a set of recommendations from the Project's Editorial Panel. It was amended on 28 December 2004 when a name for the Vocabulary was decided. The present revision is an update to acknowledge the completion of the Project and launch of the IPSV. However, the underlying policy has not changed.

## Revision History

Revision date	Previous revision date	Summary of changes
2004-11-01		First issue
2004-12-28	2004-12-28	Vocabulary renamed as "Integrated Public Sector Vocabulary"
2005-04-19	2004-12-28	Update to reflect launch of IPSV

## 1 Introduction

The IPSV was developed by merging three widely used vocabularies: the GCL, LGCL and SeamlessUK taxonomy. (The latter will be abbreviated to “SeamlessUK” in the present paper.) The work was overseen by the owners of the original vocabularies, who set up the Vocabulary Merger Project. The Project was completed and the IPSV launched in April 2005.

While the original vocabularies had much in common, they also had differences of content and of editorial policy. An Editorial Panel with user representation was established to consider the differences and agree policy to guide the merger Project.

The first edition of the present document described the issues needing to be resolved, before setting out the agreed policy. It has now been revised lightly, retaining the background on the issues while reflecting the circumstances that have changed in the intervening months.

Now that the IPSV has been launched, a great many users are having to adjust to the changes as they migrate from one of the original vocabularies to the new one. The present document should help them compare the policies and practices of the past with those required in the transition.

Editorial policies may well evolve in future, especially now that the IPSV is available for study and comment. Users are invited to post their comments on the discussion forum at <http://www.esd.org.uk/Forums/viewforum.php?f=39>. New users can register and login from <http://www.esd-toolkit.org/integratedforums/>.

Policies discussed herein include:

- Scope
- Abridged version
- Editorial standards applicable
- Admission and form of terms
- Category/Keyword distinction
- Node labels
- Hierarchical relationships
- Equivalence relationships
- Associative relationships
- Scope notes
- History notes
- Mappings to external facilities
- Other fields
- Change procedures and issues

## 2 Scope

The subject scope of the IPSV includes all that within the three original vocabularies. This scope was not reviewed critically during the Project, because of time pressures.

Advice from the Editorial Panel and from many local authorities suggests it would be helpful to expand the scope to cover subjects of internal interest, such as financial or human resource management. Feedback indicates that some extension to meet needs in Scotland, Wales and Northern Ireland would be useful too. It is expected, subject to approval from the IPSV Governing Board, that one or more future projects may expand the scope in these directions.

### 3 Abridged version

During the course of the Project the Editorial Panel and others considered a proposal to create subsets of the IPSV for particular user groups. Eventually it was decided that the creation and maintenance of subsets would be better handled locally, by people close to the needs of the users in question, than centrally.

An exception was made, however, on behalf of the very many prospective IPSV users, who are not local authorities nor direct providers of community information. An abridged version of the IPSV has been prepared, including all the terms at the higher levels of the IPSV, but excluding the detail specific to those two sub-sectors.

### 4 Editorial standards applicable

ISO 2788 (= BS 5723) applies to the IPSV, including the abridged version. The ISO 2788 rules for admission of hierarchical relationships have been relaxed somewhat, so as to control the total number of hierarchies and make them easier to browse. The rules for retention and factoring of compound terms have been interpreted liberally, in line with user context.

### 5 Admission criteria and form of terms

Both SeamlessUK and the GCL follow ISO 2788 rules for the form of terms, expressing concepts in the form of nouns, noun phrases or verbal nouns. Most LGCL terms conform too, but there are a few verbs and adjectives, such as “Expel” and “Crazy”. Where retained in the IPSV, such terms have been converted to conform with ISO 2788, for example to “Expulsion” or “Crazy people”.

The GCL avoids special characters in terms, such as %, & or quotation marks, because they often cause problems when imported into diverse systems. Even apostrophe can cause problems, although the GCL does admit a few of them, to avoid giving the appearance of sloppiness.

Names of organisations are another controversial item, partly because there are so many of them it is hard to know where to stop, and partly because it is customary to use a separate authority file for controlling them. The GCL has a very few, including Royal Air Force, European Parliament, UK Parliament, United Nations, but not any names of government departments. The LGCL has very few as preferred terms (unless you count Church of England and names of other churches) but it has some as non-preferred terms, e.g. Boundstone Adult Education Centre. SeamlessUK has a few more, for example Learning and Skills Council, Workers Educational Association and U3A.

Policy for the IPSV:

**Apply the ISO 2788 rules for form of terms.**

**Use precoordinated terms where useful to avoid ambiguity for terms which appear in different hierarchies with different meanings.**

**Avoid ampersand, preferring ‘and’ in its place. Admit apostrophes where they are essential to the meaning of the term. Avoid other punctuation marks and special characters wherever possible.**

**Provide lead-in entries for abbreviations and acronyms wherever helpful, with the full form as preferred term. Exceptionally, some abbreviations such as ‘VAT’ may be preferred if they are very widely used, unambiguous and more convenient than the expanded version; in all such cases a lead-in from the full form should be provided.**

**Avoid admitting names of organisations, unless they are of national importance and the public is likely to have exceptional interest in them, for example ‘Parliament’ or ‘Royal Air Force’**

Names of other organisations are considered to be outside the scope of the IPSV, but authority lists of them are still judged important for metadata applications. The Editorial Panel urges the Metadata Working Group and the Local e-Government Standards Body to consider developing additional lists and /or guidance for public bodies on best practice.

## 6 Category/Keyword distinction

The e-GMS current version requires publishers to designate Subject metadata terms as being either “Categories” or “Keywords”. Despite the explanation in the e-GMS, the basis of the distinction has never been clear to users in general. All preferred terms in the GCL are supposed to be used as Categories. The intention of the authors of the e-GMS was that all terms in vocabularies such as LGCL and SeamlessUK would be applied as Keywords. But the LGCL as published indicates that only some of its preferred terms are Keywords; the ones at higher levels are Categories. Confusion has been widespread.

Work is in hand in one of the Metadata Working Group committees to eliminate this confusion, perhaps by dropping the Keyword and Category refinements. This change will probably be approved soon after the IPSV is launched.

Policy for the IPSV:

**Drop any distinction between Keywords and Categories in the IPSV or its abridged version.**

## 7 Node labels

SeamlessUK is the only vocabulary of the original three to have node labels. This feature is potentially very useful for helping users navigate and select the right term. But it is almost impossible to find software capable of handling the node labels properly. Since the SeamlessUK use of node labels is not extensive, this feature has been dropped in the IPSV.

Policy for the IPSV:

**Node labels not to be used**

## 8 Hierarchical relationships

The three original vocabularies use the hierarchical relationship in similar ways. But SeamlessUK has slightly deeper hierarchies, and is monohierarchical, whereas both GCL and LGCL are polyhierarchical. Some users have reported difficulties with polyhierarchy, because systems such as Microsoft CMS are incapable of implementing it. On the other hand, polyhierarchy does offer superior navigation capabilities. There is some indication that software vendors are gradually upgrading to be able to cope with polyhierarchical structures.

Meantime, Porism has developed a method of converting LGCL outputs to monohierarchical, by designating one BT as the default and then providing an XSLT transformation on the esd-toolkit. It should be noted that the software being used to build the IPSV does not have this capability inbuilt. However, a “default broader term” can be assigned manually, and this data may be used by Porism to generate monohierarchical outputs as before.

Policy for the IPSV:

**Polyhierarchy should be adopted, but provisions made for converting the outputs to monohierarchical while user demand continues**

## 9 Equivalence relationships

All three original vocabularies use the equivalence relationship. The ratio of preferred to non-preferred terms is approximately 1:3, 1:2, 3:1 for the GCL, LGCL and SeamlessUK

respectively. Unlike the other two, the LGCL has a policy of sub-classifying the non-preferred terms as follows:

- synonyms
- acronyms and abbreviations
- misspellings
- antonyms
- search terms – described as “terms which are not directly equivalent to the preferred term but might be equivalent when compounded with another term..... Search terms may appear as non-preferred terms for one or more concepts.”

ISO 2788 does not require differentiation of the first four of these, but neither does it disapprove. It would be feasible to apply this practice to the whole of the IPSV, although it requires a little more effort to do so. Also, the classification is not always clear-cut (e.g. is “VAT returns” or “AS levels” an acronym or is it not?). However, we have to ask what purpose it serves. The only one for which a clear advantage is known is misspellings. By putting the misspellings in a group apart, it becomes possible to suppress displaying them to human users, but apply them when interpreting terms entered by the user.

Another potential issue is singular/plural variants. ISO 2788 does not exactly prohibit these, but they tend to conflict with the rules for which to prefer, and displays become tedious for human users if they are shown for every term.

Policy for the IPSV:

**ISO 2788 recommendations for equivalence relationships should be followed, without formally differentiating between types of equivalence. However, if misspellings are admitted, they should be separated from normal equivalents in such a way that implementers can choose to include or exclude them.**

The LGCL search terms are a different case, in conflict with ISO 2788. Examples in the LGCL are “Joining”, “Fees” and “Course fees”, which all point to “Adult and community education”. (Incidentally, they do not seem to point to any other preferred terms.) In the case of “Course fees”, it would be feasible to include in the IPSV entries such as:

Course fees (adult education) USE Adult and community education

Course fees (Music tuition) USE Music tuition

Course fees (Higher National Certificate courses) USE Higher National Certificate courses

In theory, the same thing could be done for “Fees” and “Joining”. But these terms apply to so many activities (Dental care, Universities, Housing associations, Neighbourhood Watch schemes, Voluntary organisations, etc) that it seems fruitless to clutter up the IPSV. People tend to think what they want to join before asking to join it.

Policy for the IPSV:

**“Search terms” should not be admitted as non-preferred terms or a type of equivalence relationship.**

**Existing “search terms” in the LGCL should be considered on their merits. If it is possible to admit them as normal non-preferred terms, perhaps with the addition of a qualifier, then this should be done. But spurious and misleading entries should be avoided.**

It should be noted that most of the existing “search terms” in the LGCL are admissible without offending ISO 2788, once a suitable qualifier is added. ( A qualifier is a word or phrase in brackets, added to disambiguate the main term.)

## 10 Associative relationships

All three original vocabularies include associative (see also) relationships between terms, the GCL more liberally than the other two. It is probable that few user organisations are implementing them at present, because implementers are mostly unfamiliar with their potential for aiding retrieval. But familiarity and imaginative exploitation are on the increase.

Policy for the IPSV:

**Continue existing policies and practice for associative relationships, aiming to achieve a proportion similar to that in the existing GCL.**

## 11 Scope notes

All three original vocabularies have some scope notes, but not enough. It is important to add more, especially for terms which have a special meaning for officials, and for terms that are being applied inconsistently.

Policy for the IPSV:

**Encourage all stakeholders to submit suggestions for scope notes.**

## 12 History notes

The LGCL is the only one to have this feature. Example HNs from the LGCL include, “Changed to category in 1.03”; “Moved to Illnesses and conditions in 1.02”. It is common in well established thesauri to have HNs in the style, “Introduced in 1998; prior to that see ‘Earlier term’”.

Until recently, there has not been much need for HNs, because few organisations have been using the lists properly, but this situation is rapidly changing. Some questions arise, however:

- For whom are the History Notes intended? For searchers, advising them on term usage for retrieval? Or for system administrators, advising them how to implement an update? Or for list editors, maintaining a record of what they did, when and why?
- Which history to include in the IPSV? The existing HNs in the LGCL will be valid only for LGCL history, not for the terms as used in the other vocabularies.

In theory it would be possible to have separate history notes for each of the three original vocabularies, and also for the different classes of user. And there is no technical problem in doing so. But it seems like overkill, and potentially confusing for editors and users. Furthermore, the changes to the LGCL during the merger phase are likely to be so extensive that the history notes will become very complex and earlier history not so useful.

Policy for the IPSV:

**A facility for adding History Notes should be adopted**

**The History Notes may be useful for searchers and/or administrators; feedback should be sought**

**The History Note slate should be wiped clean at the time the IPSV is launched, since this will effectively be the first version of a new vocabulary. However, it should begin to be used immediately after launch.**

**In addition to the History Notes in the IPSV, a separate audit trail of changes made to the IPSV should be maintained, starting from its launch date. This should include background on the reasoning behind decisions.**

The Editorial Panel noted that the transition to the IPSV would be difficult for many system administrators. While History Notes would not solve this problem, other forms of support and

assistance might be provided. The Panel urged the Project Board to consider this matter further.

## 13 Mappings to external facilities

The LGCL as currently maintained on esd-toolkit shows mappings to several other lists: currently the LGSL (previously known as PID list), the GCL and the ACL (APLAWS Category List), with LGCS planned to come shortly, and in future there could be more. The LGSL mappings especially are considered an essential service by many local authorities, and it is important that these should be maintained throughout this project and for the foreseeable future. However, maintenance of the mappings does not form part of the Vocabulary Merger Project.

Since the mappings are between LGCL concepts and items in the other lists, it is necessary to retain a note of each of the LGCL concepts in the IPSV, even when the preferred term changes its spelling or hierarchical location. For this reason, the LGCL concept IDs are being entered for each corresponding preferred term record in the IPSV. This should enable matching the LGSL and ACL items to the correct IPSV preferred terms, when the project is completed. The matching process needs to be tested as soon as possible.

Policy for the IPSV:

**Include LGCL Concept IDs for each relevant record, so the LGCL subset that is output can be correctly married up with the mapping facilities on the esd-toolkit.**

## 14 Other fields

The LGCL has an additional field, “A to Z”, containing either “Yes” or “No” to indicate whether the term in question should be included in local authority website A-Zs. The field applies to both preferred and non-preferred terms. SeamlessUK also has a field with a flag to indicate inclusion in the Essex website A-Z list. The GCL does not have anything comparable – it expects user organisations to make their own decisions about website directories.

It would be entirely feasible to include both these flags in separate fields in the IPSV, for use with the corresponding subsets. However, it is likely that the existing flags will need to be reviewed when the merger is complete, as the changes will be quite extensive.

The Editorial Panel noted that the flags for A-Z lists and shortcuts both lie outside the scope of the merger Project. However, user soundings could be taken during the course of the Project with a view to ensuring the needed facilities are still available when the merger is complete.

The LGCL also has a field indicating availability of a shortcut. It is used infrequently – for about 40 shortcuts altogether.

## 15 Change procedures and issues

### 15.1 Term deletions

The editors of all three vocabularies have a policy of avoiding deletions wherever possible. It is normally preferable to demote a term to non-preferred status, indicating the appropriate preferred term.

Published policy of the LGCL is quite explicit: “Once added, terms will never be removed from the List.” Unfortunately, however, there are some very unsuitable terms (mostly non-preferred terms) in the existing List, which will give users trouble all the time they remain. Examples include: “Locations & hours”, “Support”, “Suspend”, “Muddled”. Some of them contravene ISO 2788. System administrators may complain, but users will find the List and its subsets vastly improved if all such terms are removed. There are also some names of local

organisations which are useful only to particular authorities and should not clutter the national vocabularies

Fortunately, there are some ways of mitigating the problems of removing such terms. Firstly, some terms can be modified to become acceptable. For example, “Pupil support” is admissible, or “Financial support (students)”, or “Supported accommodation”. Secondly, the LGCL has a group of terms called “Obsolete terms”. It would be possible to relegate all the unwanted terms to this category, and suppress them from outputs where they are not required. After a suitable period, say one or two years, they could then be dropped altogether. Thirdly, the appropriate authorities could be asked to “take back” their local terms and handle them locally.

Policy for the IPSV:

**Until such time as the merger is launched, suspend the policy of embargoing deletions, while maintaining a list of deleted terms under “Obsolete terms”.**

**Use best endeavours to retain at least a modified version of the deletion candidates as non-preferred terms.**

**When the merger is complete and the IPSV launched, it will be important to reinstate the policy of never deleting a concept. Terms may be modified or moved or demoted, but as a minimum a lead-in entry should be retained from the term in its previous preferred form to the appropriate current preferred term.**

Note: at the time of launch, a list of deleted terms has not been published. There are some practical difficulties where terms have been substantially modified, in deciding whether the modification is so great that effectively the term has been deleted. In lieu of showing deletions, mappings have been built showing what current term(s) to use for each previous LGCL concept ID. Feedback is awaited on whether more should be done.

## 15.2 Approval of changes

Both the GCL and the LGCL have had procedures for receiving proposals for change, consulting users about them and giving ultimate authority to a user committee. Such procedures are vital to maintain quality and give users confidence in the products. During the merger project, however, the changes were too numerous and it was not practical to consult about each individual one. Following the launch, consultation procedures will soon be announced.