



## IPSV Maintenance Guide

IPSV Maintenance Guide.....	1
1. Introduction.....	2
1.1 Simple explanation of purpose .....	2
1.2 Background and context.....	3
2. Editorial policy .....	4
2.1 Scope .....	4
2.2 Abridged version.....	4
2.3 Editorial standards applicable.....	4
2.4 Admission criteria and form of terms .....	4
2.5 Category/Keyword distinction .....	4
2.6 Node labels.....	4
2.7 Hierarchical relationships .....	5
2.8 Equivalence relationships.....	5
2.9 Associative relationships .....	5
2.10 Scope notes.....	5
2.11 History notes.....	5
2.12 Mappings to external facilities .....	5
2.13 Term deletions.....	5
3. Infrastructure for maintenance and user involvement .....	6
4. Consultation procedure .....	6
5. Maintenance set-up and procedure.....	7
5.1 Database structure in MultiTes.....	7
5.2 Editing the database.....	8
5.3 Database outputs .....	10
6. References on thesaurus construction .....	12
7. Additional IPSV documentation.....	13

### *Revision history of this Guide*

Revision date	Previous revision date	Summary of changes
2005-12-31	2005-06-14	Provision for term deletions and retention of Concept ID and Term ID during updates
2006-03-31	2005-12-31	Minor update to align with IPSV V2
2006-08-29	2006-03-31	Minor update to align with e-GMS V3.1

## Metadata

Coverage.spatial	UK
Creator	Stella Dextre Clarke, Consultant
Date.issued	2005-06-14
Date.modified	2006-08-29
Description	Explanation of how to manage and maintain the IPSV (Integrated Public Sector Vocabulary)
Format	Text
Language	eng
Publisher	Porism Limited, London, SW9 8BJ info@porism.com
Rights.copyright	Crown copyright. This document may be used and copied without payment or licence for research, private study or for internal circulation within an organisation. This is subject to the material being reproduced accurately and not used in a misleading context. The source and copyright status must be acknowledged in any copy of whole or part of the document. Any other proposed use of the document requires a copyright licence, which is available from <a href="http://www.opsi.gov.uk">www.opsi.gov.uk</a>
Status	Version 1.3. For publication
Subject	Controlled vocabularies
Title	IPSV Maintenance Guide
Type	Instructional

### 1. Introduction

The present document is a guide to maintenance of the Integrated Public Sector Vocabulary (IPSV), a controlled vocabulary of approximately 3,000 preferred terms and 4,000 non-preferred. The IPSV serves as an encoding scheme for use with the Subject element of the [e-Government Metadata Standard](#) (e-GMS).

The IPSV aims to help with access to resources produced in the UK public and voluntary sectors, and has broad coverage of all the relevant subjects. It provides greater detail in areas that concern the daily lives of citizens and communities.

#### 1.1 Simple explanation of purpose

The function of the IPSV is to help users find things easily and reliably. Users cannot find the right items among a collection of hundreds of resources, unless these are presented so that everything on the same subject comes together. A search using a keyword such as “Jobs” typically fails to retrieve items described as “vacancies”, “recruitment opportunities”, “situations vacant”, “staff openings”, etc. To ensure that search results are reliable, it is good to label information resources with consistent names describing what they are about.

The IPSV is a guide to naming concepts consistently, so that they can be found. It needs to be used together with tools for applying the consistent names in metadata, and search engines that read the metadata.

## 1.2 Background and context

The e-GMS has always recommended the use of controlled vocabularies for Subject metadata. Back in 2002, the e-GMS offered two refinements of the Subject element, namely Category and Keyword. Subject.category was intended for terms describing the broad subject area of a resource, while Subject.keyword would accommodate detailed indexing at the level of specificity of the resource. A controlled vocabulary called the Government Category List (GCL) was mandated for use in populating Subject.category. The GCL was quite small, with just around 400 preferred terms, but had a wide scope covering all subjects relevant to the UK public sector.

The use of additional controlled vocabularies has always been optional in the e-GMS, and several have developed. The Local Government Category List (LGCL) was widely used among local authorities, for Subject.keyword as well as Subject.category, and also as a website navigation structure. A much bigger vocabulary called seamlessUK was also used in parts of the public and voluntary sectors, to support indexing in post-coordinate retrieval systems, and was applied to Subject.keyword.

With the exception of local authorities, few public sector bodies found any of the lists suitable for retrieval of the information resources produced by themselves. For this the typical organisation needed a vocabulary with specificity comparable to seamlessUK, but subject coverage appropriate to its own responsibilities. Many such thesauri and taxonomies already existed in bodies such as the Department for Education and Skills, Countryside Agency and Ministry of Defence, and the e-GMS has encouraged use of specialised vocabularies like these. At the same time, there has been an obligation to apply the GCL, and a widely adopted model was to map the specialised vocabularies to the umbrella structure of the GCL.

In 2004 the decision was taken to merge the GCL, LGCL and seamlessUK taxonomy, leading to the launch of IPSV in April 2005. The product is a tool which supports detailed indexing in the subjects of interest to local authorities and the providers of community information. In other subject areas the IPSV does not support detailed indexing but may be used for broad categorisation. It has not been designed as a web navigation structure, for any organisation.

To make things easier for organisations whose subject areas are not well covered, an abridged version of the IPSV is maintained. It is about one-fifth of the size of the full IPSV and exactly the same at the upper levels. It can be used for broad categorisation of resources. Just as with the previous GCL, any public body may choose to apply its own specialised vocabulary and map it to the abridged IPSV, this serving as an umbrella structure that is interoperable across the public sector. (This does not apply to local authorities, who are required by a directive from the Department for Communities and Local Government (DCLG) to use the full IPSV.)

As of IPSV Version 2.00 the scheme has been extended to cover 'internal-facing' subjects (since the first version was limited to 'citizen-facing' subjects). A subset called the 'Internal Vocabulary' comprises all the terms thought useful for indexing the internal information resources of public sector bodies.

To make things easier for local authorities requiring a standard web navigation structure, a tool called the Local Government Navigation List ([LGNL](#)) has been developed quite separately and is available via the [esd-toolkit](#).

When Version 3.1 of the e-GMS was released in August 2006, the Category and Keyword refinements were deprecated and IPSV replaced the GCL as the mandatory encoding scheme for Subject, using the unrefined element. The e-GMS allows organizations (other than those complying with DCLG requirements) to choose whether to use full IPSV or the abridged version, and gives examples showing the correct syntax in each case. Rigorous use of correct syntax makes it possible for a search engine to use the metadata appropriately.

## **2. Editorial policy**

The following policy statements were approved to apply to the merger project. They are shown here with some restatement to reflect that the merger is now complete.

### **2.1 Scope**

At the time of launch, the subject scope of the IPSV includes all that within the three original vocabularies. 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.

### **2.2 Abridged version**

An abridged version of the IPSV should be maintained, including all the terms at the higher levels of the IPSV, but excluding the detail specific to local authorities and the providers of community information. The abridged version should cover the full scope of the IPSV and remain suitable for broad categorisation purposes.

### **2.3 Editorial standards applicable**

The IPSV, including its abridged version, should comply with ISO 2788, the International Standard for monolingual thesauri and BS 8723, the corresponding British Standard. The ISO 2788 rules for admission of hierarchical relationships should be 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 should be interpreted liberally, in line with user context.

### **2.4 Admission criteria and form of terms**

ISO 2788 rules apply to the form of terms.

Pre-coordinated terms are admitted 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 outside the scope of the IPSV.

### **2.5 Category/Keyword distinction**

There is no distinction between Keywords and Categories in the IPSV or its abridged version or internal vocabulary subset.

### **2.6 Node labels**

Node labels are not used in the IPSV.

## **2.7 Hierarchical relationships**

The IPSV is polyhierarchical, but provisions are made for converting the outputs to monohierarchical while user demand continues.

## **2.8 Equivalence relationships**

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

The type of "Search terms" previously included in the LGCL should be admitted as non-preferred terms only if they comply with ISO 2788.

## **2.9 Associative relationships**

Associative relationships should be admitted in line with ISO 2788.

## **2.10 Scope notes**

Scope notes are considered very useful and users are encouraged to submit suggestions.

## **2.11 History notes**

A facility for adding history notes should be adopted.

History notes may be useful in different ways for searchers and/or for system 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.

## **2.12 Mappings to external facilities**

All LGCL Concept IDs should have a mapping to the corresponding IPSV term(s).

## **2.13 Term deletions**

Once a concept is admitted to IPSV it should be retained wherever possible. If the preferred term for a given concept should change, then the new preferred term should retain the Concept ID of the previously preferred term. Very occasionally a concept may need to be removed (for example if two concepts are considered so close that they need to be merged as one, or if a subject area or level of specificity is pruned right out of the Vocabulary). In such cases, the existing terms should preferably be retained as non-preferred terms.

Second in priority is to retain all terms, especially preferred terms, wherever possible. If a preferred term is changed, then preferably the previous form of it should be retained as a lead-in (non-preferred term) to the new preferred form. Non-preferred terms should be removed only when their retention would be misleading or confusing.

A list of any preferred terms removed from the IPSV should be maintained and made available to users and implementers. At the time of writing, it is shown in the introduction of the hierarchical style Word version of IPSV.

### **3. Infrastructure for maintenance and user involvement**

The IPSV is maintained by a single Editor, reporting to the IPSV Governing Board. Editorial policy is determined by an Editorial Panel, with representation from users in all parts of the UK public sector. The Editorial Panel supports the Editor by arbitrating when conflicting user demands arise.

The Editor maintains the IPSV using a package called MultiTes 2005 Pro, which was designed specifically for ISO2788-compliant thesauri. The package generates outputs in non-proprietary formats, and so the option is open to move to another package in future. But for simplicity, the present Guide will be written around the assumption of using MultiTes.

IPSV updates are issued only once or twice a year, after a batch of amendments is complete. Outputs from MultiTes are supplied to a hosting service, currently Porism Ltd. Porism loads the latest version online on the [esd-toolkit](#), and generates some additional outputs. All the outputs, together with guidance notes, are presented for downloading from the esd-toolkit. While a different hosting service could be chosen in future, the present Guide will be written around the assumption that it is Porism.

It is important to be aware that Porism generates some additional fields, not present in the MultiTes database but derived from the MultiTes outputs. These fields are described in the [Guide to IPSV Versions and Formats](#). Two of the Porism fields need care from the Editor: Concept ID and Term ID. The rationale for these fields is as follows:

- Each term in IPSV, preferred or non-preferred, should have a unique Term ID. The Term ID should remain the same regardless of any changes to the term's relationships with other terms, and regardless of changes in scope notes, history notes, mappings, inclusion in subsets, but should change if there is any change in the form of the term itself.
- Each concept in IPSV should have a unique Concept ID. The same Concept ID applies to a preferred term and all of its non-preferred terms. For preferred terms, the Concept ID will often (but not invariably) be identical to the Term ID.
- In MultiTes, the Concept ID of a preferred term will be given by its Term Record Number (TNR). The Concept ID of a non-preferred term is given by the TNR of the corresponding preferred term. Normally the Term ID of any term (preferred or non-preferred) in MultiTes is given by its own TNR. But in a small number of cases a Previous TNR (PTN) field is present. In these cases, the Previous TNR takes precedence over the current TNR, as the source of the Term ID.

Also on the esd-toolkit is a [discussion forum](#), open to any member of the public on [registration](#). Suggestions and proposals for updating the IPSV are aired on the forum, so that all users have an opportunity to comment.

### **4. Consultation procedure**

The main principles are as follows:

- All significant proposals for amendments to the IPSV are announced on the esd-toolkit discussion forum, with a minimum of three weeks for public comment before implementation proceeds. (An exception is made for minor amendments such as addition of more non-preferred terms or related term (RT) relationships, which are unlikely to upset existing users of the preferred terms. It seems sensible not to obscure the more important proposals with a flood of minor details.)
- Typically each suggestion submitted needs some research before it can be accepted. The Editor does the research, which may include correspondence with the proposer, either by email or via the forum, as appropriate.

- Where the proposal is in line with editorial policy and is not controversial, the Editor may proceed to implement it in the current update.
- Where user opinions are in conflict and/or where the Editor has reservations about the wisdom of the proposal, he circulates it to the Editorial Panel for advice. A reasonable consensus must be obtained before any such proposal is approved.

These procedures are still evolving and may be developed further. In future some meetings of the Editorial Panel may be arranged.

## 5. Maintenance set-up and procedure

### 5.1 Database structure in MultiTes

Attributes and relationships	Tag	Explanation
<i>Attributes</i>		
Preferred term		See ISO 2788
Scope note	SN	See ISO 2788
History note	HN	See ISO 2788
Subset code	SC	See Notes
Term record number	TNR	See Notes
Previous TNR	PTN	See Notes
Date of input	INP	See Notes
Date last updated	UPD	See Notes
LGCL Concept ID	LID	See Notes
Default broader term	DBT	See Notes
<i>Relationships</i>		
Corresponding non-preferred term	UF	See ISO 2788
Misspelt non-preferred term	MS	See Notes
Broader term	BT	See ISO 2788
Narrower term	NT	See ISO 2788
Related term	RT	See ISO 2788
<b>Table 1: Structure of the database record for a preferred term</b>		

#### Notes to Table 1:

**Subset codes.** Some of the terms in the IPSV are allocated to subsets. Those belonging to the abridged version have a character G (for abridGed version) in this field. Terms in the Internal Vocabulary subset have N (for iNternal). A term may belong to more than one subset.

**Term record number.** This is a sequential number assigned automatically by MultiTes each time a new term is entered. It does not change if the term is subsequently amended, for example to correct the spelling. The record number of a preferred term serves also as the unique identifier of the concept represented by that term.

**Previous TNR.** When amendments are made, sometimes they result in a term acquiring a different TNR. When this happens, the old TNR should be entered manually in this field. PTN should be filled both for the term that gets changed and the one that takes its place.

**Date of input and Date last updated.** These dates are assigned automatically by MultiTes each time a record is added or edited. It is important to set the format as YYYY-MM-DD, in line with international standards.

**LGCL Concept ID.** At the time when IPSV was created, it was important to provide mappings from preferred terms in the LGCL. This field may hold one or more LGCL IDs, separated by

semi-colons and no spaces. At the time of writing, it is not clear whether updating of these IDs will be needed in future.

**Default broader term.** Some systems are unable to import polyhierarchical data. When an IPSV term has more than one Broader Term, it is necessary to designate one of them as the default, so that the IPSV can be imported as a monohierarchy.

**Misspelt non-preferred term.** These are included to provide extra entry points in the event that a user searches with a misspelt keyword.

Attributes and relationships	Tag	Explanation
<i>Attributes</i>		
Non-preferred term		See ISO 2788
History note	HN	See ISO 2788
Subset code	SC	See Notes
Term record number	TNR	See Notes
Previous TNR	PTN	See Notes
Date of input	INP	See Notes
Date last updated	UPD	See Notes
<i>Relationships</i>		
Corresponding preferred term	USE	See ISO 2788
Corresponding preferred term with normal spelling	NS	See Notes
<b>Table 2: Structure of the database record for a non-preferred term</b>		

### Notes to Table 2:

**Subset codes.** Some of the terms in the IPSV are allocated to subsets. Those belonging to the abridged version have a character G (for abridGed version) in this field. Terms in the Internal Vocabulary subset have N (for iNternal). A term may belong to more than one subset.

**Term record number.** This is a sequential number assigned automatically by MultiTes each time a new term is entered. It does not change if the term is subsequently amended, for example to correct the spelling. Non-preferred terms do not have concept IDs directly assigned to them, but the record number of the corresponding preferred term serves also as the concept ID for the non-preferred term.

**Previous TNR.** When amendments are made, sometimes they result in a term acquiring a different TNR. When this happens, the old TNR should be entered manually in this field. PTN should be filled both for the term that gets changed and the one that takes its place.

**Date of input and Date last updated.** These dates are assigned automatically by MultiTes each time a record is added or edited. It is important to set the format as YYYY-MM-DD, in line with international standards.

**Corresponding preferred term with normal spelling.** This is filled only in cases where the non-preferred term is a misspelling of the preferred term.

## 5.2 Editing the database

The general principles of thesaurus maintenance are described in ISO 2788<sup>1</sup> and in the manual by Aitchison et al<sup>2</sup>. The practicalities of using MultiTes are described in a series of tutorials<sup>5</sup> from the MultiTes website. Advice in the present document is limited to the additional special needs of IPSV.

Requests for IPSV amendments can come from anybody, and one cannot assume that the proposer understands the constraints of ISO 2788, or the multiple meanings a candidate term may have, or the need for choosing names and structure that work generally throughout the UK. Therefore it is important to research every suggestion carefully, and challenge it (in a tactful way!) where appropriate. Here is a short checklist:

- Check all proposals carefully. Very often the originator has overlooked an existing term which duplicates or conflicts with the proposal. It is important to avoid duplication and to minimise the amount of overlap between preferred terms.
- Is the proposed term really needed? Could it be lumped with an existing preferred term, perhaps with the addition of a non-preferred term?
- Is the new concept expressed appropriately? Does the proposed term need a qualifier to clarify the intention?
- If the term is accepted, will the scope of existing preferred terms need adjusting? Will some non-preferred terms need changing? Should new non-preferred terms be provided? Will a scope note be required?
- Where does the new term best fit in the hierarchies? Should it go in more than one place? Will an RT link be needed from an existing term?
- Would the new term add unacceptably to the number of terms under a broader term? Or require an extra level to be added?
- If the proposal is for a structural change, how will this affect existing users and systems?

User-friendliness is a key factor in choosing terms. Very often the meta-tagging job is given to people who are not experts in the subject matter. They cannot be expected to make fine distinctions between technical terms. Therefore the vocabulary must be generally easy-to-read and unambiguous across different organisations.

Cross-sectoral perspective is vital. Citizens enquiring about a public service may not know whether it comes from local authorities, regional or central government, much less which particular department is responsible. The structure and terminology of the IPSV may occasionally coincide with Civil Service usage, but must not rely on any prior knowledge of traditional organisational boundaries. Jargon expressions are often useful as non-preferred terms, but preferred terms should be widely comprehensible and unambiguous.

Unlike most thesauri, IPSV needs to have a very limited number of top terms, to ease navigation for users who want to “drill down” from the top. To achieve this, it is necessary to relax the strict ISO 2788 rules for hierarchical relationships. When adding new preferred terms to the database, it is important always to enter them as narrower terms of an existing preferred term.

MultiTees allows many alternative ways of making the same editorial amendment. It is important to choose a way that allows an existing concept to retain the same TNR (and hence Concept ID), even if the preferred term for it changes or new relationships are added. As a second priority each term needs to retain the same Term ID, even if its TNR gets changed in the effort to leave the TNR with the preferred term for the concept. (And every term must end up with a unique Term ID.) This supports continuity for users, when updates are released, but can only be achieved by taking great care while editing.

For example, it is not a good idea to delete an existing preferred term and then add a separate new one, if the basic concept is the same, because the new one will be given a completely different TNR. As another example, suppose you want to swap an existing preferred term and one of its non-preferred terms. The best method is to amend the spelling of each of them, without deleting either of them. By this means, the TNRs will be swapped, but the conceptID will remain the same. Both UPDs will record the date of the change. After this is done, it is necessary to make entries for the previous TNR of each term, so that the terms can retain their old Term IDs in the outputs generated by Porism.

Some of the non-preferred terms in the IPSV database are misspellings. These are handled differently from normal preferred terms, so that they can be suppressed from database outputs if necessary. They should be entered using the MS tag instead of UF. The reciprocal tag is NS, for ‘Normal spelling’. The Editorial Policy section gives criteria for entering misspellings.

In the run-up to an update, the editing of all the terms, scope/history notes and relationships should be completed before some final checks on default broader terms (DBTs) and the subset codes are done, as follows.

For the DBTs, first check that all the records with an existing DBT entry really do have more than one broader term. Where there is only one broader term, remove the DBT entry. Secondly, check that all preferred terms with more than one broader term do have a DBT entry. Check also that the entry is valid, i.e. it must match one of the broader terms *exactly*. Make corrections as necessary. The whole process is cumbersome because MultiTes has no inbuilt mechanism for such an unusual requirement.

For each subset it is necessary to ensure that the set of records with the corresponding code is coherent, compliant with ISO 2788, and is based on the existing top level terms, without breaks in the hierarchy. Step one is to be satisfied that users are happy with the current allocation of codes. (At the time of writing, there has been little feedback on this question. In future there might be a need for cutting back or expanding the size of either subset, and the codes should be adjusted as necessary.) With the vast bulk of the codes in place, the most important check is that every preferred term with a G code should have at least one broader term with a G code, unless it is a top term. If necessary, assign a code to any intermediate terms needed to bridge the gap. (If a G-coded term has more than one Broader term, it is only necessary that one of these has a G code.) Next, for each preferred term with a G code, check that all its non-preferred terms also have G codes. Finally, check that every non-preferred term with a G code corresponds to a preferred term with a G code. Then follow the same procedure for N codes.

### 5.3 Database outputs

When editing is complete, an output in XML is sent to Porism. This output should contain all the fields mentioned. (MultiTes offers some additional fields: Status (STA), Type (TYP), Approval date (APP) and Not valid date (NVD). These fields are not used at present and should not be included in outputs.)

As well as loading the database for online display, Porism generates some additional outputs in formats familiar to ex-users of the LGCL. These are documented in the [Guide to IPSV Versions and Formats](#). Porism does not prepare the additional outputs described below, which are generated directly from MultiTes and made available for downloading from the esd-toolkit.

#### Full IPSV Index - ISO 2788 style - Word format

This format offers a human-readable version in the conventional layout of a thesaurus compliant with ISO 2788. Generate from MultiTes as follows:

- Select Report generator
- Select all terms
- Deselect terms with NS relationship
- Select these relationships: SN, USE, UF, BT, NT, RT, HN
- Output as a text file
- Edit the text file with Word:
  - Convert to the typeface required by IPSV Board (previously Times New Roman but Arial may be preferred in future)
  - Use 'find and replace' to expand the tags and present in italics
  - Add title and introductory paragraphs in a different typeface
  - Add page numbers and a header with version number and release date
- Save as a Word file

#### Full IPSV in plain text, ISO 2788 style

This machine-readable format is for users with their own thesaurus management software, who can import it directly. (MultiTes users will find this exceptionally simple; others may need to edit the file lightly, depending on their preferred import format.) Generate as follows:

- Select Report generator
- Select all terms
- Select these relationships: TNR, SN, USE, UF, NS, MS, BT, NT, RT, DBT, LID, HN, SC, INP, UPD
- Output as a text file

### **Full IPSV - comma-delimited ASCII text**

This machine-readable format is for users who wish to import it into a database or spreadsheet. The CSV output generated by Porism is for the same type of user, and at the time of writing it is not clear whether both of these formats are needed. Until such time as feedback indicates this format is redundant, generate as follows:

- Select Report generator
- Select all terms
- Select these relationships: TNR, SN, USE, UF, NS, MS, BT, NT, RT, DBT, LID, HN, SC, INP, UPD
- Output as a comma-delimited file

### **Full IPSV in HTML**

This format is very convenient for loading on a PC, where it can be searched without special software and without going online. Some users keep the HTML version open while meta-tagging, so they can easily browse, select terms and copy-and-paste them into the meta-tagging application. MultiTes generates a full set of HTML pages, one for every term in the database plus a few index pages. Generate as follows:

- First check the HTML folder in your MT2005 directory is empty
- Select TopTerm Report
- Select all terms
- Output as HTML
- Next select Report generator
- Select all terms
- Deselect terms with NS relationship
- Select relationships: SN, USE, UF, BT, NT, RT, HN
- Select HTML output
- Leave HTML parameters as they are
- Disable search engines – unless there are reasons for including in future
- Adjust “more parameters” to show the IPSV title and subtitle
- Output the files. They will be output to the same HTML folder as before.
- Check the headers on the main index and top term pages, and edit the HTML source data as necessary.

Be sure to zip up all the HTML pages before sending them anywhere.

### **Abridged IPSV - Hierarchical style - Word format**

This format shows preferred terms only, in hierarchical style. It was used for Version 1, and is included here for completeness, but may not be needed in future since Porism is now generating a similar output in esd-standard style. Generate from MultiTes as follows:

- Select TopTerm Report
- Select all terms with Subject Category G
- Output as a text file
- Next select Report generator
- Select top terms only
- Select only the NT relationship
- Output as a second text file
- Go back to Report generator
- Select top terms only
- Output as a third text file
- Edit the first text file with Word:
  - Convert to the typeface required by IPSV Board (previously Times New Roman but Arial may be preferred in future)
  - Use ‘find and replace’ to increase the indentation space
  - Embolden all the top terms
  - Convert to two columns, taking care with wrapping of any long lines
- Edit the second text file with Word:
  - Convert to the typeface required by IPSV Board
  - Strip out the NT tags
  - Use ‘find and replace’ to adjust indentation space if necessary
  - Embolden all the top terms
  - Convert to two columns, taking care with wrapping of any long lines

Edit the third text file with Word:  
Convert to the typeface required by IPSV Board  
Embolden all the terms, adjust text size and spacing  
Assemble the three files into one document, following model of previous version  
Add title and introductory paragraphs in a different typeface  
Add page numbers and a header with version number and release date  
Save as a Word file

### **Abridged IPSV Index - ISO 2788 style - Word format**

This format offers a human-readable abridged version in the conventional layout of a thesaurus compliant with ISO 2788. Generate from MultiTees as follows:

Select Report generator  
Select all terms with G subset code  
Deselect terms with NS relationship  
Select these relationships: SN, USE, UF, BT, NT, RT, HN  
Output as a text file  
Edit the text file with Word:  
Convert to the typeface required by IPSV Board (previously Times New Roman but Arial may be preferred in future)  
Use 'find and replace' to expand the tags and present in italics  
Add title and introductory paragraphs in a different typeface  
Add page numbers and a header with abridged, version number, release date  
Save as a Word file

### **Abridged IPSV in plain text - ISO 2788 style**

This machine-readable format is for users of the abridged version with their own thesaurus management software, who can import it directly. (MultiTees users will find this exceptionally simple; others may need to edit the file lightly, depending on their preferred import format.)

Generate as follows:

Select Report generator  
Select all terms with G subset code  
Deselect terms with NS relationship  
Select these relationships: TNR, SN, USE, UF, BT, NT, RT, HN, INP, UPD  
Output as a text file

## **6. References on thesaurus construction**

1. ISO 2788-1986 Documentation - Guidelines for the establishment and development of monolingual thesauri. Geneva: International Organization for Standardization; 1986.
2. BS 8723-1:2005 Structured vocabularies for information retrieval. Guide. Definitions, symbols and abbreviations. Published 1 November 2005. ISBN 0 580 46798 8
3. BS 8723-2:2005 Structured vocabularies for information retrieval. Guide. Thesauri. Publication date: 1 November 2005 ISBN: 0 580 46799 6
4. Aitchison, Jean; Gilchrist, Alan, and Bawden, David. Thesaurus construction and use: a practical manual. 4th ed. London: Aslib; 2000. 218 pp.
5. Craven, Timothy C. Thesaurus construction: an introductory tutorial [Web Page]. Available at: <http://instruct.uwo.ca/gplis/677/thesaur/main00.htm#contents>.
6. Dextre Clarke, Stella G. Thesaural relationships. Bean, Carol A and Green, Rebecca, Eds. Relationships in the Organization of Knowledge. Dordrecht: Kluwer; 2001; pp. 37-52.
7. Lesson No.1 - Getting Started with MultiTees 2005 Pro. This and three more 'Lessons' may be downloaded from <http://www.multitees.com/download.htm>
8. Will, Leonard. Thesaurus principles and practice. Available at: <http://www.willpowerinfo.co.uk/thesprin.htm>

## 7. Additional IPSV documentation

Name of document	Comment
<a href="#">IPSV FAQs</a>	The easiest place to start
<a href="#">Guide to Meta-tagging with IPSV</a>	Advice for webmasters and authors of electronic resources when entering metadata
<a href="#">Automatic categorisation for IPSV users</a>	Tips before you implement automatic categorisation
<a href="#">Implementing IPSV to your own advantage</a>	Hints for getting more out of IPSV, and adapting it to your own needs
<a href="#">IPSV Guidance Notes</a>	Especially useful for local authority users
<a href="#">Design/selection criteria for software used to handle controlled vocabularies</a>	Helps with choosing software for any part of the implementation
<a href="#">IPSV Editorial Policy</a>	Includes discussion of the issues leading to development of the Policies now in force
<a href="#">Which IPSV? A guide to the versions and formats available</a>	Full description of the options: online display, downloadable files, full or abridged, machine-readable or for human eyes
<a href="#">e-Government Metadata Standard (e-GMS)</a>	Full details of all the metadata elements needed for interoperability in the public sector

Note: Comments on this Guide should be posted on the IPSV [discussion forum](#)