

UK Archival Thesaurus (UKAT):  
Construction and Editing Methodology

| Version | Date       | Author        | Summary   |
|---------|------------|---------------|---|
| 3.0     | 11/07/2003 | Phil Carlisle | Third draft following comments by management board      |
| 4.0     | 18/07/2003 | Phil Carlisle | Fourth draft incorporating comments from advisory panel |
| 5.0     | 18/07/2003 | Phil Carlisle | Five draft incorporating comments from Peter Garrod     |
| 6.0     | 19/08/2003 | Phil Carlisle | Sixth draft incorporating comments from Louise Craven   |

|  |    |
|--|----|
| <b>CONTENTS</b>  | 2  |
| Executive Summary                                      | 3  |
| 1. Introduction  | 5  |
| 1.1 UNESCO thesaurus                                   | 5  |
| 1.2 The NDAD thesaurus                                 | 6  |
| 1.3 Sources for UKAT terms                             | 6  |
| 2. Definition and structure of the thesaurus           | 7  |
| 2.1 Scope and coverage                                 | 7  |
| 2.2 Structure of the thesaurus                         | 9  |
| 2.2.1 Term   | 9  |
| 2.2.1.1 Singular and plural                            | 9  |
| 2.2.1.2 Definitions and scope notes                    | 12 |
| 2.2.1.3 History Notes                                  | 14 |
| 2.2.1.4 Punctuation and Capitalization                 | 14 |
| 2.2.1.5 Abbreviations and Acronyms                     | 18 |
| 2.2.1.6 Spelling and Language                          | 18 |
| 2.2.2 Relationships                                    | 20 |
| 2.2.2.1 Equivalence                                    | 20 |
| 2.2.2.2 Hierarchical                                   | 21 |
| 2.2.2.3 Associative                                    | 26 |
| 3. Editing methodology                                 | 29 |
| 3.1 Identify the requirement                           | 29 |
| 3.2 Define the scope                                   | 29 |
| 3.3 Term selection                                     | 29 |
| 3.4 Defining equivalencies                             | 30 |
| 3.5 Structuring hierarchies                            | 30 |
| 3.6 Relating terms                                     | 31 |
| 3.7 Initial data import and software testing           | 31 |
| 3.8 Editorial evaluation                               | 31 |
| 3.9 Formal evaluation                                  | 32 |
| 3.10 Web publication                                   | 32 |
| 3.11 Publicity   | 33 |
| 3.12 Ongoing development and candidate term submission | 33 |
| 4. Conclusions and Recommendations                     | 34 |
| 5. Bibliography and sources                            | 35 |

## Executive Summary

This document intends to set out the construction techniques and the editorial processes which will be involved in the development of the UK Archival Thesaurus (UKAT).

As well as outlining the details of the work involved, it makes a number of recommendations. These recommendations reflect feedback from the Management Board; from the Advisory Panel meetings on 6 May 2003 and 15 July 2003; and from individual members of the Advisory Panel.

The recommendations set out in this document are:

- UKAT should be developed as a separate thesaurus to the *UNESCO Thesaurus* (Section 2.1) with a preference for UNESCO terms being retained.
- UKAT should be developed as a polyhierarchical thesaurus (Section 2.2.2.2). Initially work should focus on the development of polyhierarchies which are needed for the inclusion of non-UNESCO terms.
- UKAT should conform to the British and ISO standards with regards to the construction of the thesaurus. (Section 2 as a whole)
- The terms derived from UNESCO should retain the form, whether singular or plural, that they already have. This will ensure that any future mapping between UNESCO and UKAT will be simplified. (Section 2.2.1.1)
- Any new terms submitted by contributing archives and projects should be edited to ensure they conform to the project's standards on singular and plural usage. (Section 2.2.1.1). A guidance note should be produced and included on the website for prospective submissions, to demonstrate the rules regarding the use of singular and plural forms, as well as other requirements for the submission of terms.
- Unique entities or 'classes of one' should be included in the thesaurus, although wholesale use of them should be discouraged. However, personal, corporate and place names will not be included and any unique entities which are added must be useful to the wider archive community. (Section 2.2.1.1)
- The project will provide scope notes for all new terms submitted to the thesaurus, and for those UNESCO terms and non-UNESCO terms created prior to UKAT which are ambiguous. (Section 2.2.1.2)
- A field should be provided for in any software used by the project to include details of changes to a term including its provenance and any editorial changes since submission. This information could be included in the alphabetical display. (Section 2.2.1.3). Although the guidelines allow for the inclusion of a History Notes, to differentiate between a field which details the development of the term and a historical definition this field should be called the Authority Note

- Punctuation marks should be avoided. (Section 2.2.1.4)
- The existing use of initial capitalization in the UNESCO Thesaurus should be retained to ensure future interoperability with the *UNESCO Thesaurus* (Section 2.2.1.4)
- The project should retain the present structure at microthesaurus level but must be willing to extend this structure and where necessary add new microthesauri. (Section 2.2.2.2)
- The thesaurus display should use BT/NT for the display. (Section 2.2.2.2)

# 1. Introduction

The UK Archival Thesaurus project has been funded by the Heritage Lottery Fund under the *Your Heritage* programme. We intend to develop a thesaurus that will provide a standardized, controlled vocabulary for subject indexing in the UK archive sector. This document is intended to provide an overview of the procedures involved in the construction and editing of thesauri. Due to the complexities and resource-hungry nature of thesaurus construction, it was decided to base UKAT on an existing standard rather than reinventing the wheel. We chose the *UNESCO Thesaurus* because it was already being used by a number of archives and archive projects, and generally conformed to British and international standards on thesauri.

## 1.1 UNESCO Thesaurus

The *UNESCO Thesaurus* includes terms derived mainly from UNESCO documents and the *OECD Macrothesaurus*. These have been grouped into seven categories or field names.

These are:

1. Education
2. Science
3. Culture
4. Social and human sciences
5. Information and communication
6. Politics, law and economics
7. Countries and country groupings

The thesaurus is predominantly in English but does include French and Spanish linguistic equivalents. The thesaurus contains:

- 87 microthesauri (each falling into one of the seven categories)
- 4261 descriptors (Preferred Terms)
- 4049 hierarchical relationships (Broader and Narrower Terms)
- 5771 associative relationships (Related Terms)

The English version also includes 2341 non-descriptors (Non-Preferred Terms)<sup>1</sup>.

Of the 6602 terms within the thesaurus only 601 (9%) have been provided with scope notes.

The structure of the thesaurus is predominantly monohierarchical, although the last field *Countries and country groupings* does allow terms to be placed polyhierarchically.

---

<sup>1</sup> The number of non-descriptors differs as each term in any language may have different synonyms. There are currently no French or Spanish non-descriptors in the thesaurus however it has been indicated that UNESCO intend to develop the thesaurus into a truly multi-lingual thesaurus.

Although the *UNESCO Thesaurus* is used widely across the sector and generally accepted as an excellent resource for subject indexing, it is increasingly being recognized that it is inadequate for indexing collections at the specific levels of detail required by archivists. This has led to many archives adding to, or amending, the thesaurus for their own needs.

## 1.2 The NDAD Thesaurus

The National Digital Archive of Datasets uses a subject index adapted from the *UNESCO Thesaurus*. Although it is a subject index rather than a true thesaurus it does contain thesaural elements, and retains much of the structure of the *UNESCO Thesaurus* with regards to its use of the hierarchical, equivalent and associative relationships.

The approach taken with the NDAD Thesaurus has been to only incorporate those terms (approx. 600) from the *UNESCO Thesaurus* that have been used to index the datasets held by NDAD.

These terms have been augmented with the addition of terms needed for indexing but not present in UNESCO (approx. 300).

## 1.3 Sources for UKAT terms

UKAT intends to take the *UNESCO Thesaurus* as its starting point with the addition of terms from several archival projects, all of whom have adopted UNESCO as a standard for subject indexing.

These projects are:

- CASBAH
- GASHE
- AIM25
- A2A
- Archives Hub<sup>2</sup>

Between them these projects account for hundreds of individual archives from both across the country and the sector as a whole. Terms provided by NDAD and derived from the National Archives' PROCAT subject index will also be dealt with at an early stage.

The terms provided by these projects will be supplemented, as the UKAT project progresses, with those submitted by other archive projects, individual repositories, and interested bodies.

---

<sup>2</sup> The Archives Hub allows contributors to submit terms taken from either UNESCO or LCSH (but not amendments to either scheme). UKAT will focus on incorporating those terms from LCSH which Hub contributors have submitted, which do not have UKAT or UNESCO equivalents but which are deemed useful to the wider archive community.

## 2. Definition and structure of the thesaurus

### 2.1 Scope and coverage

The development of the UK Archival Thesaurus is intended to fulfil the needs of the UK archive sector and provide a common standard for subject indexing, enabling users to search a wide variety of archives without the need for them to be familiar with a range of subject indices.

UKAT will cover the subject areas already extant in the *UNESCO Thesaurus*, with the exception of the *Countries and country groupings* microthesaurus, and will be expanded to include terms from other subject indexes and classification schemes, such as LCSH, which have been used for indexing by the archive partners.

UKAT not only intends to gather terms in use in major archives and archival projects, it is also committed to incorporating terms which reflect the histories and experiences of groups which have traditionally been under-represented among archive users, to encourage their participation in the archival heritage.

As part of the development of UKAT we will also look at areas which may be useful for subject indexing but which have not previously been addressed, one example being that relating to landholdings. This is an area which has been identified by several repositories but which is insufficiently catered for by existing thesauri or subject indexes in use in the archive community. UKAT will explore the possibility of creating microthesauri using terms submitted by the participating members for given subject areas. This will include identifying thesauri from non-archive sources which could be incorporated into UKAT.

With the inclusion of these additional terms, it is likely that the structure of the thesaurus will diverge considerably from that of the UNESCO. This will be particularly true as it has been agreed that UKAT should be developed as a polyhierarchical thesaurus. This divergence is unavoidable. Although any new terms added to future editions of UNESCO could be incorporated into UKAT, it would be impossible to mirror the structure. As such UKAT will become, to all intents and purposes, a stand alone thesaurus.

As a general rule, only those terms which are deemed to be useful for indexing by the community as a whole will be included. That is not to say that terms which are relevant to only one archive will not be included, but rather that a term needs to be seen in the broader context of information retrieval.

For instance, the *UNESCO Thesaurus* contains the term *databases* and has narrower terms which generic types such as *bibliographic databases*. The number of narrower terms has been increased by NDAD to include the specific names of individual databases such as *Judicial Statistics database*. As this is a very specific database, and it is unlikely that several archives would wish to index records with this term, it would not be included in UKAT.

The issue of the status of UNESCO preferred terms was raised at the Advisory Panel meeting held on 15<sup>th</sup> July 2003. It was agreed that the majority of UNESCO terms would be retained as preferred terms but with the caveat that some non-preferred terms may be upgraded depending upon the user community. Such terms will be treated on a case-by-case basis. However it is not envisaged that all UNESCO terms will be revisited in this regard and the recommendation was to concentrate on the non-UNESCO terms which are being added with UNESCO terms being dealt with if time permits.

**UKAT Recommendation:**

As a general rule, UKAT will seek to avoid incorporating 'classes of one' and terms which appear to be relevant to only one repository, while recognizing that some terms in these categories may have an overriding usefulness that leads to their incorporation.

This issue is particularly relevant to the *Countries and country groupings* microthesaurus.

Whether UKAT should include place names was discussed at the first meeting of the advisory panel. The general consensus, confirmed at the second meeting, was that maintaining a place name component would unnecessarily increase the complexity of the thesaurus and duplicate work which has already been undertaken in great detail by others, such as the Getty Vocabulary Programme's *Thesaurus of Geographic Names*<sup>3</sup> and the Great Britain Historical GIS Project (<http://www.gbhis.org/>).

It is therefore recommended that UKAT only incorporate the first six microthesauri from UNESCO. The issue of divergence with UNESCO is one which needs considerable thought. Although developing UKAT as a stand alone thesaurus is the recommended option, care must be taken to ensure that the two thesauri map across to each other. This will aid data exchange. Any changes made to the preference of an existing UNESCO term must also take this mapping into account. At no point should a UNESCO term be deleted completely from the thesaurus, but a term may be made a non-preferred term.

---

<sup>3</sup> Available online at <http://www.getty.edu/research/tools/vocabulary/tgn/>

## 2.2 Structure of the thesaurus

The UK Archival Thesaurus, will conform to the British and ISO standards for thesaurus construction. As such it will contain the following information.

- Term
- Scope Note
- Equivalent relationships
- Hierarchical relationships
- Associative relationships

The first phase in the construction of any thesaurus is the collection of terms.

### 2.2.1 Term

As already mentioned, for the purposes of UKAT most of the terms which will be included will already exist either in the *UNESCO Thesaurus* or the subject indexes currently in use by the archives community. As a result it would be difficult, not to say pointless, to start again and define a set of rules which may not apply to these pre-existing terms.

In both the ISO 2788, and its British equivalent BS 5723 *Guidelines for the establishment and development of monolingual thesauri* a term is defined as being the “representation of a concept”. A term usually exists as a single word but compound terms, made up of two or more words do occur. Where compound terms do exist they should always appear in natural language order except when they are given a parenthetical qualifier (see section 2.2.1.4: Parentheses).

For example:

**British Trees** *rather than* **Trees, British**

However **Trees (British)** may be used

#### 2.2.1.1 Singular or plural

The form in which a term appears is also discussed in the standards. Whether to use the singular or plural tends to be one of the lingering debates. Whereas it is seen to be acceptable to use the singular, if working in the French or German language, in English the plural form is preferred. However certain communities, among them the museums community, favour the use of the singular form, and the American equivalent standard ANSI/NISO Z39.19 allows this exception. The issue, as to whether to use the singular or plural, stems from the question as to how a user will query a database.

For instance if a visitor to a museum is interested in clocks then they may ask the curator “how many *clocks* do you have in your collection?” On the other hand when the curator accessions an item and adds a record to the database he or she will usually think in the singular about an individual item, ie Question: “What am I holding?” Answer: “A clock”.

If the curator has indexed the record with CLOCK then the user will not retrieve the record using a search for CLOCKS. Some search engines may truncate the term being queried, and this will obviously work for any term which forms a regular plural, however what happens if on a trip to the zoo our user is interested in knowing how many MICE are being looked after.

One solution is to include the singular (or plural) term as a non-preferred term. This way a user will be guided towards the correct term to use in their search. This approach is already used within the *Getty Art and Architecture Thesaurus*, with the plural form taking preference. In fact the AAT, an English language thesaurus, has gone one stage further and includes the British English spelling as an alternative for the American - very handy if searching the database for products made from aluminium!

The *UNESCO Thesaurus*, like the AAT, uses the plural form except in those instances where the meaning would be lost or unusable. This particularly applies to the *Country and country groupings* microthesaurus where it would be extremely unlikely that a user would submit a query asking how many AFRICAS there are in the world.

As *UNESCO* is already using the plural, and the majority of UKAT terms will be derived from it, it would be ridiculous to revert to the use of the singular form and, indeed, within the timeframe of the project, impossible.

That being said consideration must still be given to this area, particularly when providing guidance on the submission of candidate terms. The standards divide terms into three broad categories, and to a certain extent these affect the choice of singular or plural.

These are:

- **Concrete entities**
- **Abstract concepts**
- **Unique entities** (expressed as proper nouns)

### **Concrete entities**

These can be sub-divided into

- count nouns – subject to the question “how many” – expressed in the plural  
eg. BOOKS, ARCHIVES, PIGS
- non-count nouns –subject to the question “how much” – expressed in the singular  
eg. CHEESE, BEER, SOIL

However, as with most rules there are exceptions.

Parts of the body are expressed in the plural when more than one part are found

For example:

EARS and TEETH

But

NOSE and PELVIS

Also, as previously mentioned, the ANSI standard allows for the use of the singular in museum catalogues.

For non-count nouns the rule is more flexible. It allows exceptions where the user community regards a term as being a class with more than one member. In the above example in a CAMRA database there may be terms in a BEERS hierarchy, such as PORTERS, STOUTS, BITTERS, LAGERS etc.

### **Abstract concepts**

These are usually expressed in the singular. Abstract concepts include

- belief systems - BUDDHISM
- activities - HIKING
- disciplines - ARCHAEOLOGY
- properties - POLITENESS
- abstract entities and phenomena – DISCRIMINATION

As with non-count nouns, abstract concepts can be expressed in the plural where they form a class with more than one member, for example, BELIEF SYSTEMS.

### **Unique entities**

These are defined as those terms whether concrete or abstract which form a class of one. They are always expressed in the singular as proper nouns and usually fall into one of the following categories:

- places - TRAFALGAR SQUARE, NOTTINGHAM
- organizations – UNESCO, ENGLISH HERITAGE,
- periods, dates and events – PALAEOLITHIC, BASTILLE DAY, SECOND WORLD WAR
- people – WINSTON CHURCHILL, JOAN OF ARC

As previously mentioned in section 2.1, the UNESCO *Countries and country groupings* microthesaurus will not be used to provide terms for UKAT. It would similarly be difficult to incorporate personal names and corporate names, because of the question of conformity to NCA Rules. At the Advisory Panel meeting on 15 July, there appeared to be a general consensus that personal, place and corporate names should not be included in UKAT.

Consideration will be given to including other types of unique entities within the thesaurus, where the term is seen to be useful to the wider archive community and not just a single repository.

### **UKAT Recommendation;**

The terms derived from UNESCO should retain the form, whether singular or plural, that they already have. This will ensure that any future mapping between UNESCO and UKAT will be simplified.

Any new terms submitted by the archives and projects involved should be edited using the above rules as a guide. A guidance note should be produced and included on the website for prospective submissions, to demonstrate the rules regarding the use of singular and plural forms.

Personal, corporate and place names should not be included in UKAT, though other types of unique entities may be included where they are of general usefulness

#### **2.2.1.2 Definitions and scope notes**

As with all reference works, a thesaurus is only as good as the information contained within it. For any thesaurus to be useful to the lay person, as well as the specialist, each concept or term contained in it must be defined in context. Where specialist terminology exists it needs to be made intelligible to the majority of users. The mechanism by which this can be done is by the use of scope notes.

Scope notes are not intended to be definitive definitions, however some element of definition should always be included and, where it is necessary to define the term within the context of the thesaurus, this can often be expanded upon.

As well as providing a definition, scope notes may provide additional information for the user and indexer.

This can include:

- Restrictions on the meaning

For example:

##### **Data Processing**

SN: Systematic handling, manipulation, and computation of data, always involving the use of computers. Do not confuse with “information processing”

- The range of topics covered by the term

For example:

##### **Petroleum products**

SN: Includes light oils, gas oils, fuel oils, greases, etc.

- Any specific indexing instructions

For example:

##### **Statistics**

SN: Use when referring to the science of statistics. For collections of statistics, use “statistical data”, or more specific descriptor, e.g. “educational statistics”, etc.

If the scope note directs the indexer to use additional terms it is useful if the information is reciprocal. However this is not obligatory.

Where terms exist which are now deemed to be outdated or derogatory preference will be given to the modern equivalent in current usage. Where necessary the non-preferred term can be provided with a scope note of its own explaining the historical context for the term. This is particularly important where terms which were used in the past have since become unacceptable in modern usage.

In recent years, particularly when dealing with specialist thesauri, it has become the norm to provide a scope note for every term. Although not mandatory it is a good rule of thumb that, resources permitting, should be adhered to. The more information the user can be provided with the more useful the thesaurus will become for both indexing and retrieval.

**UKAT Recommendation:**

The project should provide scope notes for all new terms submitted to the thesaurus. For terms added prior to the start of the project, including those from UNESCO, AIM25, A2A etc, the creation of scope notes will not be mandatory although where there exists potential ambiguity as to the meaning of a term, a scope note should be provided. Although it would also be advantageous to provide scope notes for those UNESCO terms which are presently undefined, it is recognized that this would not be achievable within the timeframe of the project. It is also recognized that it may not always be feasible to retrospectively create scope notes for the substantial sets of terms which have already been created by repositories and projects.

### **2.2.1.3 History notes**

A history note may form part of the scope note for a term, but it is often better to include an additional field within the thesaurus to record any information relating to the editing of a term. It is intended to provide information on the origination of the term as well as any details with regard to its evolution within the thesaurus.

For example:

**COX'S ORANGE PIPPIN**

**UF** Orange Pippin

**SN** A variety of apple

**BT** APPLE

**MT** FRUIT

**HN** Originally submitted 2001 by the Guild of Apple Growers as Orange Pippin. Added to the thesaurus as non-preferred term for apple. 2002 upgraded to preferred term. 2003 amended to COX'S ORANGE PIPPIN with non-preferred term of Orange Pippin.

After discussion with the advisory panel it was decided that the field should be renamed Authority Note as the term History Note was ambiguous and could mislead users.

Additionally any changes should be automatically recorded through an auditing facility in the chosen software.

#### **UKAT Recommendation:**

A History Note field should be provided for in any software used by the project and could be included in the alphabetical display. The thesaurus software should also allow the editor and management board to track changes to the thesaurus.

### **2.2.1.4 Punctuation and Capitalization**

Although the ISO and British standard does not specify any specific rules on either punctuation or capitalization both need to be considered.

#### **Punctuation**

The overriding factor behind the creation of a thesaurus is the desire to establish a common vocabulary that can be used without let and hindrance by a wide user community. If a thesaurus deals purely with single word terms it is easy to standardize their form and ensure that punctuation is kept to a minimum. Wherever punctuation cannot be avoided without losing the meaning of a term it should be retained. However where it is included for aesthetic value or convenience it should be excluded.

The *UNESCO Thesaurus* makes use of punctuation in compound terms. Where possible, for the purposes of UKAT, these terms should be separated into their component terms.

For example:

*information/library administration* becomes

information administration

library administration

Where such a term has narrower terms, these should automatically become the narrower term of both broader terms (in a polyhierarchical approach). However in certain cases it may be necessary to select which of the two new terms is the most appropriate broader term.

For examples:

*Information/library personnel* includes as narrower terms *Archive personnel*, *Information scientists*, *Librarians* and *Library technicians* as narrower terms.

It is clear that *Information scientists* should be made a narrower term of *Information personnel* and *Librarians*, a narrower term of *Library Personnel* but which broader term should be used for *Archive personnel*?

One solution would be to make *Archive personnel* a narrower term of *Information personnel*.

Another solution might be to increase the prominence of *Archive personnel* and put it on the same level as the other two terms.

The latter solution would make more sense as the narrower terms of all three would be on the same hierarchical level and as such would reflect reality.

## Hyphens

The standards recommend that hyphenated words should be avoided except where removal of the hyphen would render the term ambiguous or meaningless to the user. As a general rule the hyphen should be removed from compound terms but retained for those terms containing letter-word or number word combinations.

For example:

Compound Terms

PART-TIME WORK becomes PART TIME WORK

IN-SERVICE TRAINING becomes IN SERVICE TRAINING

Letter-word/number-word combinations

X-RAY FLUORESCENCE

U-BOAT

5-HYDROXYINDOLE

UNESCO includes at least two instances of hyphens in terms referring to language groups, eg. Indo-european languages.

### **Ampersand (&)**

The ampersand symbol '&' should only be retained where it forms part of a trademark or proper name. Otherwise it should be written out in full.

For example:

Marks & Spencer

but

Laurel and Hardy

### **Parentheses ( )**

Parentheses are used in thesauri to denote qualifiers for homographs and homonyms.

For example:

Bridge (Card Game)

Bridge (Structure)

As a result the use of parentheses within the term itself should be avoided. The exceptions to this rule relates to scientific compounds where the parentheses may form part of the term.

For example:

N (2 FLUORENYL) ACETAMIDE.

For the purposes of UKAT it is unlikely that scientific compounds would be included and as such UKAT terms should not include parentheses.

### **Apostrophes**

The ISO and British standards do not include guidance on the use of apostrophes with regards to the possessive case. However the ANSI/NISO standard recommends they should be retained.

The *UNESCO Thesaurus* allows the use of the apostrophe although this appears to be limited to the French language entries where terms such as *Groupe d'âge* retain the original French spelling.

### **UKAT Recommendation:**

All of the above punctuation marks may create retrieval difficulties for both the indexing and retrieval of terms. Therefore as a general rule all forms of punctuation should be avoided.

### **Capitalization**

The form which a term takes when displayed is often decided by the choice of software or the personal preference of the team responsible for constructing the thesaurus. The ISO and British standards do not cover capitalization at all but the ANSI/NISO standard recommends that capitals

should only be used for the initial letter of proper names, trade names and those acronyms which have passed into common usage such as LASER and AIDS.

The ANSI standard also recommends that capital letters should be retained where they appear in unusual positions within words such as corporate or product names, as this indicates to the user the correct orthography of a term, and helps to distinguish between common nouns and proper names.

For example:

exeGesIS SDM Ltd

In some thesauri in use in the UK, capitalization is used to indicate the status of a term. In the National Monument Record's thesauri, preferred terms are depicted in CAPITAL LETTERS whilst non-preferred terms appear in Initial Capitals.

For example:

PILLBOX

UF Pickett Hamilton Fort

In the thesauri made available by the **mda**. Capitalization has been removed entirely except in relation to the rules set out in the ANSI standard. This was a conscious decision born out of common practice in the indexing of terms in museum catalogues.<sup>4</sup>

The *UNESCO Thesaurus* uses Initial Capitalization for the first word of all the terms and retains capital letters for classes of one.

For example:

Cost reduction

but

Costa Rica

**UKAT Recommendation:**

UKAT should use initial capitalization as it is already in use in UNESCO, and this will ensure future interoperability between the two thesauri.

---

<sup>4</sup> see *Guidelines for Constructing a Museum Object Name Thesaurus* (Holm 1993, mda)

### 2.2.1.5 Abbreviations and Acronyms

Abbreviations and acronyms should only be included as preferred terms where they have become so well established that they are used in everyday language and the full term is rarely used. The full term should be included as a non-preferred term to help guide the user.

For example

Acquired Immunodeficiency Syndrome

USE AIDS

Where an acronym may be ambiguous the full form should be preferred.

For example:

WWF USE World Wildlife Fund

WWF USE World Wrestling Federation

### 2.2.1.6 Spelling and Language

As a general rule UKAT will conform to standard British English spelling. The *Oxford English Dictionary* in its various forms will be used as a source for the spelling of terms. Exceptions to this rule are:

- loan words which have become current in everyday spoken and written English eg. Entente cordiale
- French and Spanish equivalents where they already exist in UNESCO

Where a term for the same concept exists in more than one language English will be given preference except where the loan word is more widely accepted in the domain of the thesaurus.

Thus:

World weariness UF Weltschmeltz

But

Palaeolithic UF Old Stone Age

#### **The use of –ize or –ise and their variants**

The *UNESCO Thesaurus* for the most part uses the –ize form where it occurs and this is the form favoured in the *OED*, although –ise is given as an alternative spelling. However some words should only be spelt with –ise.

#### **Natural language order**

The standards recommend the use of natural language order for compound terms except where parenthetical qualifiers are included. The *UNESCO Thesaurus* conforms to this rule however one example has been identified where the term preferred by UNESCO seems at odds with the commonly accepted usage.

The term *War prisoners* occurs as a linguistic equivalent of *Prisonnier de guerre* and *Prisionero de guerra*. The term in common usage, is of course, Prisoners of war and yet this is not even given as a non-preferred term. As the standards allow for the inclusion of prepositional noun phrases in natural language order this seems rather a strange choice to make.

## 2.2.2 Relationships

A thesaurus is primarily concerned with suggesting preferred terms for use in describing for example, types of archive items, and establishing relationships between them.

There are three basic relationships within a thesaurus. These are:

- the equivalence relationship
- the hierarchical relationship
- the associative relationship

In combination with the scope note, these relationships define a term and place it in context within the scope of the thesaurus.

To construct a thesaurus, these relationships are applied to each term.

### 2.2.2.1 Equivalence relationship

This is the first relationship to be decided. A term can be 'preferred' or 'non-preferred'; however, it is the preferred term that will be used in the hierarchies and for indexing.

Non-preferred terms appear in the alphabetical listing of the thesaurus but are not placed in the hierarchy. In the alphabetical listing a non-preferred term will be cross-referenced to its preferred term using the convention 'USE'

For example:

Quantum mechanics  
USE Quantum theory

This cross-referencing is reciprocated in the display of the preferred term using the convention 'UF' or Used For.

For example:

Quantum theory  
UF Quantum mechanics

A non-preferred term is defined as a term that has the equivalent meaning to the preferred term but is not used for indexing. This might be because the term is:

- a Synonym  
eg. Car and Automobile
- a Quasi-Synonym  
eg. House and Cottage
- a Foreign term or international language variation  
eg. Pavement and Sidewalk
- a Regional Term  
eg. Alley and Ginnell
- an Archaic Term

eg. Inuit and Eskimos

In addition the equivalence relationship may include:

- plurals,
- terms in unnatural word order
- linguistic equivalents

These last three are optional and will obviously increase the number of terms included in the thesaurus. However these terms will not be used for indexing but simply as lead-in, or guide terms, for the preferred terms and as such careful consideration should be given to their necessity.

### **2.2.2.2 Hierarchical relationship**

Having established the equivalence relationship the second stage is to group the preferred terms into hierarchies. They are first gathered into conceptual groups and then within each group the terms are further divided into levels going from the broadest type of term to the narrowest and most specific type of term. In the alphabetical listing the hierarchical relationship is displayed using the following conventions:

BT (ie. Broader Term) preceding the superordinate term

NT (ie Narrower Term) preceding the subordinate term.

The hierarchical relationship is the distinguishing feature of a thesaurus. It defines the levels of superordination and subordination between:

- a class and its members (the generic relationship)
- a whole and its parts (the whole-part relationship)
- and classes of one (the instance relationship)

#### **The generic relationship**

This is the most common form of hierarchical relationship. This relationship is defined as being the link between a class and its members. It is governed by the 'all-some' rule demonstrated below.



## INDUSTRIAL ARCHAEOLOGY

- Hierarchical social structure  
eg. ARMY

DIVISION

REGIMENT

In the alphabetical listing the whole-part relationship may be indicated using the conventions:

BTP (Broader Term Partitive)

NTP (Narrower Term Partitive)

### **The instance relationship**

This relationship covers those terms which are deemed to be classes of one. Often these types of term are not included in the thesaurus but held within an additional field. However where they do occur their relationship with the broader term is an example of the instance relationship.

For example:

Record Offices

NT Nottinghamshire Record Office

Wiltshire Record Office

In the alphabetical listing the instance relationship may be indicated using the conventions:

BTI (Broader Term instance)

NTI (Narrower Term instance)

### **UKAT Recommendation:**

The Management Team and Advisory Panel were asked to consider which approach to take with regards to the use of BT/BTG, NT/NTG in the display of the thesaurus.

If BTG, BTP and BTI and their narrower term equivalents were to be used, then the definitions, with worked examples, would need to be included on the website.

For the sake of simplicity it is recommended that UKAT use BT/NT. This will also make the display appear less cluttered.

### **Monohierarchical and Polyhierarchical relationships**

It is becoming increasingly common for thesauri to be constructed using a polyhierarchical structure. A polyhierarchical approach can more accurately reflect reality in that it allows terms which occur, correctly, under more than one category to be placed in more than one hierarchy.

For example;

In a thesaurus concerned with food and drink it would be equally valid for the following entries to occur.

**DAIRY PRODUCTS**

NT MILK  
 CHEESE  
 CREAM

**DRINKS**

NT MILK  
 COFFEE  
 TEA

**MILK**

BT DAIRY PRODUCTS  
 DRINKS

This flexibility is the main advantage that a polyhierarchy has over a monohierarchy, particularly with regards to search and retrieval. In a monohierarchical thesaurus the user has to understand the thinking behind the structure of the thesaurus to be able to find a term. This is as true of even the most established thesaurus.

As Leonard Will points out even the “Art and architecture thesaurus (AAT), regrettably, allows a term to have only one broader term, and I think that this is a serious drawback to its usefulness as an information retrieval tool”.<sup>5</sup>

With the exception of the *Country and country groupings* microthesaurus, the current form of the *UNESCO Thesaurus* has a monohierarchical structure. As well as presenting difficulties for the search and retrieval process this can lead to some anomalies.

One such example is the term *Civil war*.

At present this is a narrower term of *Political Movements* in the *Politics and Government* microthesaurus. Although civil wars are inevitably associated with political, social or religious differences, they are by their very definition still wars. Unfortunately due to the monohierarchical structure this generic relationship with the broader term *War* is ignored, and is not even included as a related term, even though it might be argued that the natural place to begin searching for terms relating to war would be with the broadest term.

Any user searching a database and interested in those records associated with war in all its forms must not only be aware of this structure, but once they have identified the terms they must then have to perform several searches in order to retrieve all records which are relevant. This would not be necessary if *Civil war* was subordinate to both *Political movements* and *War*, as it would only require a single search on the term *War* (including its narrower terms).

---

<sup>5</sup> See <http://www.willpowerinfo.co.uk/thesprin.htm>

Other examples of this, taken from the *UNESCO Thesaurus*, include:

- *Drama* as a narrower term of *Literary forms and genres* and only a related term of *Theatre*. Confused even further by having *Dramatic arts* as a non-preferred term for *Theatre*
- *Animal genetics* as a narrower term of *Zoology* and not *Genetics*, which itself has the narrower term of *Human genetics*
- *Animal diseases* as a narrower term of *Zoology*, and related to *Pathology* which has the narrower term of *Diseases*
- *River basins* as a narrower term of *Rivers* and only related to *Basins*
- *Satellite broadcasting* as a narrower term of *Broadcasting technology*, whereas both *Radio* and *Television* are narrower terms of *Broadcasting* even though they are both examples of *Broadcasting technology*
- *Houses* and *Palaces* are both only related terms of *Buildings*

The advantage of a polyhierarchical thesaurus stems from its flexibility. However it does have implications for the way in which a term is used within a database, and how the uniqueness of the term is recorded<sup>6</sup>.

To fully benefit from a polyhierarchical structure, the database used for indexing must treat the term *itself* as unique rather than its position in the hierarchy. This will ensure that when the term is used to retrieve records the search mechanism will search for all those records indexed with the term itself.

In this way if, in the above example, *Civil war* is a narrower term of both *Political movements* and *War* any search using either of the broader terms (which also asks for any narrower terms) should retrieve all records indexed with the term regardless of the context placed on the term by the indexer.

For example:

If the indexer has a document concerned with the Anarchist movement in Barcelona during the Spanish Civil War the context of the term *Civil war* relates in this instance to the *Political movements* broader term. However, a researcher interested in documents relating to the military aspect of the Spanish Civil War may still be interested in this document. Therefore if that user enters the term *Civil war* they will want to retrieve it, even if they are accessing the term through its broader term *War* and even if they ultimately disregard the record as being irrelevant for the particular enquiry.

#### **UKAT Recommendation:**

The project should retain the present structure at microthesaurus level but must be willing to extend this structure and where necessary add new microthesauri. The project should develop UKAT as a polyhierarchical thesaurus to ensure the flexibility of retrieval and the correction of any

---

<sup>6</sup> For an example of the table structure of a thesaurus database see Appendix 1.

anomalous terms. However, given the timeframe of the project work will focus on the development of polyhierarchies which will be useful for the inclusion of the newly submitted non-UNESCO terms. The UNESCO hierarchies will only be modified where anomalies, as in the above examples, occur. This approach should not present a problem for the databases used by the majority of the UKAT user base, as the major archive software suppliers (including Adlib Information Systems' archive module and DS Ltd's CALM for Archives) support polyhierarchical thesauri. As most archives will be using one or other of these packages, the issue of database structure is really only relevant to institutions/projects using 'bespoke' database software (e.g. NDAD!).

### **2.2.2.3 Associative relationship**

This relationship allows terms which are linked conceptually but not hierarchically or as equivalents. The standards define these “related terms” as being terms which are “mentally associated to such an extent that the link between them should be made explicit in the thesaurus, on the grounds that it would reveal alternative terms which might be used for indexing and retrieval”.

This relationship is always reciprocal so if Term A is related to Term B then Term B is automatically related to Term A. In the alphabetical listing the associative relationship is indicated using the conventions:

RT (Related Term) preceding the associated term thus:

Archives

RT     Records management

It is not usually necessary to relate terms belong to the same category (or microthesaurus) as they are deemed to be related by a common broader term which when used for retrieval would bring back results for the sibling terms.

The exception to this rule occurs when the sibling terms have overlapping meanings and may be loosely used in place of one another, even though both may be precisely defined and as a result are not equivalents. The example which the standards use is ‘ships and boats’ which are both narrower terms of ‘vessels’, but are somewhat interchangeable in their use.

### **Terms in separate hierarchies**

Where terms exist in separate hierarchies they can easily be related according to the scope of the thesaurus but care must be taken in particular with the whole-part associative relationship.

As mentioned above, except in the four instances of hierarchical whole-part relationships, the whole-part relationship is deemed to be associative. This is easy to see when the terms are dealing with concrete entities. In a specialized thesaurus dealing purely with buildings the relationship may take the form of the whole-part hierarchical relationship.

For example:

BUILDING

NT WINDOW

However in a more general thesaurus where the relationship may not be unique to a pair of terms, then the associative relationship takes preference.

For example:

BUILDING

RT WINDOW

CAR

RT WINDOW

AEROPLANE

RT WINDOW

Other than those governed by the whole-part relationship, there are numerous types of related term and some of them are given in the standards. Some of these include:

- Disciplines and the objects studied  
SEISMOLOGY  
RT EARTHQUAKES
- An operation or process and its agent or instrument  
EXCAVATION  
RT TROWEL
- Occupations and the person employed  
ARCHIVE SCIENCE  
RT ARCHIVIST
- Action and its product  
BREWING  
RT BEER
- A concept and its opposite  
BIRTH  
RT DEATH

Although it is useful to be aware of these various types of related term, it is only necessary for the thesaurus compiler to establish that the relationship is associative. This can be done by ensuring that the term is neither an equivalent term (where the equivalent relationship would take precedence) or hierarchically linked (where the hierarchical relationship would take precedence).

**UKAT Recommendation:**

UKAT should retain those associative relationships which already exist in the UNESCO Thesaurus, with new associative relationships being formed for each new term added where appropriate.

Where an associative relationship with a possible broader term exists due to the monohierarchical structure of UNESCO (see the *Houses* and *Palaces* example), these will be reviewed and revised.

## **3 Editing Methodology**

This section will focus on the processes involved in the creation and editing of the thesaurus. It aims to set out a step by step procedure which can be used as a template for the editorial functions as well outlining the procedures for dealing with candidate terms.

### **3.1 Identify the requirement**

The first step when contemplating the development of a new thesaurus is identifying the requirement. The construction of a thesaurus from scratch is an extremely labour and resource hungry task and as a result it should not be entered into lightly. Careful consideration needs to be given as to whether pre-existing thesauri could be used or modified to meet the requirements. The requirement for a more detailed, archive-focused, terminology has been identified to expand upon the *UNESCO Thesaurus* and this requirement will be met by UKAT<sup>7</sup>. UKAT will use pre-existing controlled vocabularies (including locally developed vocabularies) as sources for terms, within an overall structure which will be based on UNESCO but subject to modification (see pp. 24-25).

### **3.2 Define the scope**

Once the requirement for a thesaurus has been identified it is then necessary to define the scope of the thesaurus. This includes the level of specificity of the thesaurus (ie. the level of detail of the terms included), the subject areas covered and excluded, the display and the layout as well as any system requirements. This document, and the recommendations in it, is intended to help the Management Team and Advisory Panel to agree the scope of the thesaurus (see section 2.1 Scope and Coverage).

### **3.3 Term selection**

As the need for a greater level of detail, with regards to the terms required for indexing by archives, has already been identified, the process of collecting terms for inclusion in the thesaurus has already begun. As noted in section 1.3 the major sources will be:

- Pre-existing thesauri and controlled vocabularies
- Classification schemes

It is also usual practice to derive information for terms from the following:

- Encyclopaedias and reference books
- Dictionaries
- Journals and abstract journals dealing with the subject area
- Other publications dealing covering the subject

<sup>7</sup> See Subject Indexing in UK Archives: The Case for a UK Archival Thesaurus

This does not mean that new terms will be sought from these, but rather that they will be used to supplement the understanding of the term, and may be used to help the editor understand the correct hierarchical position of a term. They allow the construction of a fully rounded thesaurus as opposed to one built in isolation from its subject matter.

UKAT will use the *UNESCO Thesaurus* as the primary source for its terms, as well as the underlying structure (subject to modifications). This will be augmented using candidate terms supplied by the archives and projects involved.

Initially data will be imported using datasets supplied by major archival projects and some archives (NDAD, National Archives). The terms derived from these will be treated as candidate terms, although it is anticipated that the majority will be accepted without any objection.

During the later phases of the project it will be possible for projects, archives and individuals to submit candidate terms electronically, either by email or an online form, as well as by traditional means. The submission of large numbers of terms by an archive may also be done electronically with an import/export routine. Guidance will be provided to contributors on acceptable data formats.

### **3.4 Defining equivalencies**

As the terms to be included will be obtained from various sources, it will be necessary to establish whether any terms submitted are synonyms. Establishing which terms will be preferred will be determined by several factors, although as a rule preference will be given to those terms which are in most common use.

The rules regarding the equivalence relationship, as defined in section 2.2.2.1, will be used to determine the preferred term where synonyms, quasi-synonyms, regional, foreign and archaic terms occur.

### **3.5 Structuring hierarchies**

Having established which terms will be preferred it is then necessary to place the terms into hierarchies.

The initial stage of hierarchical construction will be to assign each term to the relevant microthesaurus. These microthesauri will be further developed with the inclusion of additional terms where they are deemed necessary to improve the structure with regards to search and retrieval.

It is anticipated that extensive revision of the hierarchical structure will be required: in particular, the development of UKAT as a polyhierarchical thesaurus will necessitate this.

### **3.6 Relating terms**

This is often the most difficult and time consuming part of thesaurus construction and as a result it is usually the last thing to be undertaken. As has been shown in section 2.2.2.2 there is the additional issue of UNESCO related terms which are actually polyhierarchical broader terms. Depending on the number of terms in the thesaurus and the level of specificity, the number of associative relationships can run into the hundreds. Each related term needs to be examined individually and the reciprocity ensured.

### **3.7 Initial data imports and software testing**

Once the initial imports of data have been received from the UKAT partners, a detailed review of the thesaurus will be undertaken. This will be a detailed look at the main hierarchies and the relationships between terms. It will also cover the forms of the terms, paying particular attention to spelling and, where it exists, consistent use of punctuation.

It is hoped that the software which we are going to use to construct the thesaurus will be available by the end of July at the latest. Testing of the software will take place to ensure that it fully meets the requirements of the project. The software will be supplied with the latest version of the UNESCO Thesaurus, and the other datasets will be imported into it using an import/export routine.

### **3.8 Editorial evaluation**

Once all the initial data has been provided and imported into the software, the editing process will begin. This will be an ongoing process which will continue until the end of the project. It will involve changes to the hierarchical structure, alterations to spelling and punctuation, evaluation of candidate terms, and the dissemination of these changes to the advisory panel for approval. It is hoped that a fortnightly bulletin, detailing changes to the thesaurus can be circulated to the members of the advisory panel by email. This will allow any comments/objections to be taken into account before the change is published to the website.

Where more information is required on a term it may be necessary to designate a member of the advisory panel to research the term and feedback to the editor. This will quite often involve very little work and the advisory panel should not see this as an onerous task<sup>8</sup>.

Once the initial contributions have been dealt with, the thesaurus will 'go live' on the web. This will allow the user base to become involved in the development of UKAT and provide continuous feedback. It is hoped that users will review the development of the thesaurus and evaluate the success, or failure, of the product. The feedback received through the website will always be acted upon and where possible the reviewer contacted to discuss their comments. This will ensure that the reviewers feel they are actively involved.

---

<sup>8</sup> This is the role of 'subject specialist' which is ominously referred to in the candidate term flowchart in appendix 4.

As part of this initial revision, a test should be carried out using the thesaurus to index documents. This will help to highlight any glaring errors in subject coverage as well as testing the structure of the relationships.

### **3.9 Formal evaluation**

Towards the end of the project (in months 11-14) both the thesaurus and the website will undergo formal evaluation. This will look at both the technical aspects of the project and the content of the thesaurus itself.

Any alterations recommended by this evaluation with regards to the software will be actioned in months 14-15 and any content changes will be undertaken in months 15-16.

### **3.10 Web publication**

At present the thesaurus will be published online and will also be made available in digital format to archives and the major suppliers of archive software. The website will include an introduction detailing the structure and scope of the thesaurus as well as an alphabetical and hierarchical listing of the terms. There should also be an acknowledgements page and a bibliography or list of sources which details the main sources for the terms. It is also common practice to include a comments form which can be used to submit candidate terms<sup>9</sup>

At the end of the project a copy of the thesaurus should be deposited with Aslib in the UK and with the international clearinghouse based in the University of Toronto, Canada.

This holds copies of digital and printed thesauri produced in the English language as well as multilingual thesauri with English language sections.

The address of the clearinghouse is:

Subject Analysis Systems Collection  
Faculty of Library Information Science  
University of Toronto  
140 St George Street  
Toronto  
Ontario M55 1A1  
Canada

As well as publishing the thesaurus online, it will also be made available, free of charge, to software suppliers for incorporation into their archives packages. As most archives use proprietary off the shelf packages this will ensure that the UKAT will be available almost immediately.

### **3.11 Publicity**

---

<sup>9</sup> For examples of printed and online submission forms see Appendices 2 and 3

Throughout the life of the project it is intended to publicize the development of UKAT. Publicity will take the form of articles, talks at archive-related conferences and seminars (ongoing for the life of the project) and the production of leaflets aimed at the archive community (months 6 and 7). It is also intended to hold two seminars to demonstrate the work of the project to those groups who are traditionally under-represented. These are scheduled to take place in month 9 and month 16.

### **3.12 Ongoing development and candidate term submission**

After the completion of the project it is likely that candidate terms will continue to be submitted. As a result a mechanism will need to be established to administer the future development of the thesaurus. This could take the form of a working party or an email discussion forum. Each proposed candidate term should be processed within an agreed timeframe. The procedures for long-term maintenance should be agreed by the UKAT panel, prior to completion of the project, and set out on the thesaurus website.

The likely procedures involved will be<sup>10</sup>:

1. Candidate term is submitted via the online comments form.
2. The term is reviewed to ensure that it is not already present in the thesaurus.
3. A list of the latest candidate terms is produced and distributed for consideration to the discussion forum.
4. The members of the forum provide feedback and a decision is made as to the status of the term, whether preferred or non-preferred, and its position in the hierarchy.
5. Any terms on which a decision has not been made may be passed on to a designated member of the forum for further consideration and then re-circulated.
6. The decisions of the forum members are made available.

---

<sup>10</sup> See also Appendix 4: Candidate term flowchart.

## 4 Conclusions and recommendations

The development of any thesaurus is labour intensive. The adaptation of an existing thesaurus is intended to save time but can often involve as much, if not more work than the creation of a wholly new thesaurus.

Given the short timeframe of the project, and the resources available, it is essential that work is focussed on those areas which are deemed to be most important for the end user, namely the increase in terms and the functionality with regards to the use of the thesaurus for search and retrieval.

The recommendations set out in this document are:

- UKAT should be developed as a separate thesaurus to the *UNESCO Thesaurus* (Section 2.1) with a preference for UNESCO terms being retained.
- UKAT should be developed as a polyhierarchical thesaurus (Section 2.2.2.2). Initially work should focus on the development of polyhierarchies which are needed for the inclusion of non-UNESCO terms.
- UKAT should conform to the British and ISO standards with regards to the construction of the thesaurus. (Section 2 as a whole)
- The terms derived from UNESCO should retain the form, whether singular or plural, that they already have. This will ensure that any future mapping between UNESCO and UKAT will be simplified. (Section 2.2.1.1)
- Any new terms submitted by contributing archives and projects should be edited to ensure they conform to the project's standards on singular and plural usage. (Section 2.2.1.1). A guidance note should be produced and included on the website for prospective submissions, to demonstrate the rules regarding the use of singular and plural forms, as well as other requirements for the submission of terms.
- Unique entities or 'classes of one' should be included in the thesaurus, although wholesale use of them should be discouraged. However, personal, corporate and place names will not be included and any unique entities which are added must be useful to the wider archive community. (Section 2.2.1.1)
- The project will provide scope notes for all new terms submitted to the thesaurus, and for those UNESCO terms and non-UNESCO terms created prior to UKAT which are ambiguous. (Section 2.2.1.2)
- A field should be provided for in any software used by the project to include details of changes to a term including its provenance and any editorial changes since submission. This information could be included in the alphabetical display. (Section 2.2.1.3). Although the guidelines allow for the inclusion of a History Notes, to differentiate between a field

which details the development of the term and a historical definition this field should be called the Authority Note

- Punctuation marks should be avoided. (Section 2.2.1.4)
- The existing use of initial capitalization in the UNESCO Thesaurus should be retained to ensure future interoperability with the *UNESCO Thesaurus* (Section 2.2.1.4)
- The project should retain the present structure at microthesaurus level but must be willing to extend this structure and where necessary add new microthesauri. (Section 2.2.2.2)
- The thesaurus display should use BT/NT for the display. (Section 2.2.2.2)

## 5. Bibliography and Sources

No paper which discusses thesaurus construction can be written without recourse to the first two sources given here.

Aitchison J., Gilchrist A., Bawden D. (1997): *Thesaurus Construction and use: a practical manual. 3<sup>rd</sup> Edition 1997* (London: Aslib)

British Standards Institution (BSI)(1987): *BS 5723: Guidelines for the establishment and development of monolingual thesauri, 2<sup>nd</sup> Edition.* (London: BSI)

English Heritage (1999): National Monuments Record Thesauri: Introduction  
Available online at <http://www.english-heritage.org.uk/thesaurus/newuser.htm>

Holm, S. (1993). *Guidelines for Constructing a Museum Object Name Thesaurus.* (Cambridge: mda)

International Organization for Standardization (ISO) (1986): *ISO 2788: Guidelines for the establishment and development of monolingual thesauri, 2<sup>nd</sup> Edition* (Geneva: ISO)

National Information Standards Organization (1994) *ANSI/NISO Z39.19: 1993. Guidelines for the Construction, Format and Management of Monolingual Thesauri.* (Bethesda: NISO Press)  
Also available online at <http://www.niso.org/standards/resources/Z39-19.html>

UNESCO (1995): *UNESCO Thesaurus* (Paris: UNESCO)

Weiner E. S. C., Delahunty A. (1994) *The Oxford Guide to English Usage Second Edition* (Oxford: OUP)

Willpower Information (1996). *Thesaurus principles and practice.*  
Available online at <http://www.willpower.demon.co.uk/thesprin.htm>.

## Appendix 1: English Heritage AMIE Table Structure

The English Heritage AMIE database and the Thesaurus Management module incorporated within it is a bespoke ORACLE database. As well as the ability to query the database through the windows graphical user interface forms it is also possible to query and edit the data using SQL Plus.

The Thesaurus Management module consists of four tables:

- thesaurus\_terms
- thesaurus\_term\_preferences
- thesaurus\_term\_uses
- thesaurus\_term\_relations

It also includes a table called classification\_groups which holds details about the individual thesauri.

**thesaurus\_terms** This table contains the basic information about each term in the thesaurus.

***the\_te\_uid*** is its unique number in this table and is the link to the preference table.

***term*** is the actual term and is the link to the uses table.

***Index\_term*** indicates if the term should be index or not.

***status*** indicates if a term is preferred (P), non-preferred (N) or candidate (C)

***scope note*** is the description of the term.

***cla\_gr\_uid*** denotes which classification the term is in and is the link to classification\_groups. The inclusion of this field enables the same term to be included in several thesauri without the use of homographs as the term + cla\_gr\_uid creates the unique identifier for the term

**thesaurus\_term\_uses** This table contains the relationships between terms and their use/positions in the hierarchies.

***th\_t\_u\_uid*** is the unique number within this table and is used to link terms within the table and it is used to link to the relation table.

***term*** is used to link to the terms table.

***cla\_gr\_uid*** denotes which thesaurus the term is in.

***broad\_term\_u\_uid*** holds the th\_t\_u\_uid of the term immediately above in the hierarchy, ie its broader term.

***top\_term\_u\_uid*** holds the highest th\_t\_u\_uid for a term, ie its class term.

**thesaurus\_term\_preferences** This table links terms from the term table together. This is to give preferred and non-preferred terms.

**the\_te\_uid\_1** is the the\_te\_uid of the non-preferred term and links to the the\_te\_uid column of the term table.

**the\_te\_uid\_2** is the the\_te\_uid of the preferred term that the\_te\_uid\_1 is to be linked to. the\_te\_uid\_2 links to the the\_te\_uid column in the term table.

**thesaurus\_term\_relations** This table links terms in the hierarchy together as related terms.

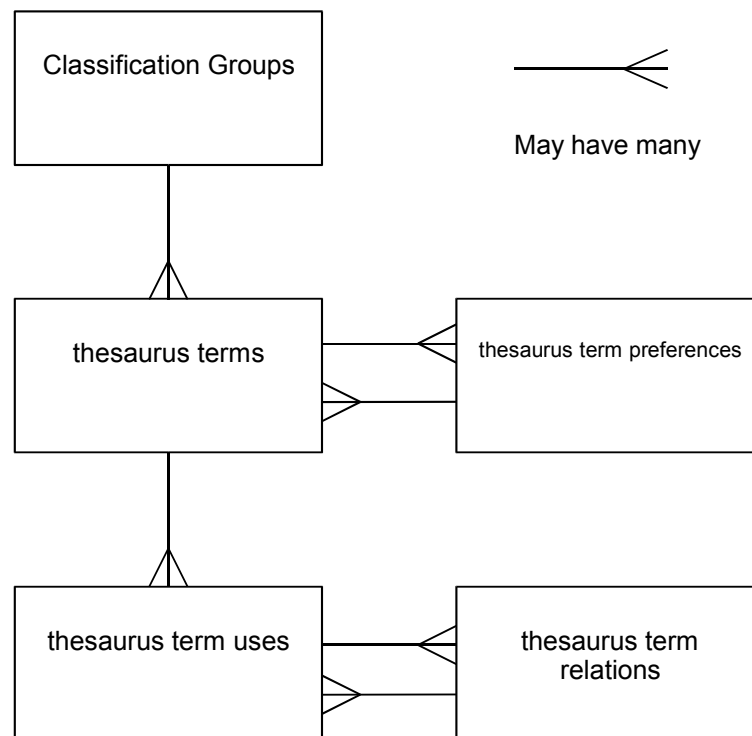
**th\_t\_u\_uid\_1** is the th\_t\_u\_uid from the thesaurus\_term\_uses table of the first term to be related

**th\_t\_u\_uid\_2** is the th\_t\_u\_uid from the uses table that are to be connected. The relationship is always reciprocal.

### classification\_groups

This stores information about the thesaurus or authority list. It is linked to the thesaurus tables through the cla\_gr\_uid.

### Physical data model of the AMIE thesaurus tables.



The thesaurus terms are linked to the recording modules of AMIE through the the\_te\_uid which appears in the monument classifications table as term\_uid.



|                          |                    |                      |         |
|--------------------------|--------------------|----------------------|---------|
| <b>Date Received</b>     |                    | <b>Date Reviewed</b> |         |
| <b>Decision</b>          | Accepted           | Rejected             | Pending |
| <b>Action taken</b>      | Added to thesaurus | Moved                | Deleted |
| <b>Reason for action</b> |                    |                      |         |
| <b>Date actioned</b>     |                    | <b>Signed:</b>       |         |

### Appendix 3: Example of an online candidate term submission form

The screenshot shows a web browser window titled "Heritage thesauri - comments - Microsoft Internet Explorer". The address bar displays "http://www.english-heritage.org.uk/thesaurus/comments.htm". The main content area is titled "Propose a candidate term:" and includes a note: "Indicates that you must enter something in order to submit a term successfully." The form consists of several fields:

- Thesaurus:** A dropdown menu with the text "[Choose the thesaurus you wish to submit a candidate term to]" and two visible options: "Monument Types" and "Archaeological Objects".
- Candidate Term:** A text input field.
- Source of Term:** A text input field.
- Published references:** A large text area with a vertical scrollbar.
- Scope Note:** A large text area with a vertical scrollbar.

At the bottom of the form is a button labeled "Send This Mail Message Now".

The above figure shows part of the online submission form for the National Monuments Record Thesauri. The user fills in their contact details along with any comments and the details about the candidate term and then simply submits the form. This is forwarded, in the form of an email, directly to the thesaurus editor.

Although there is no registration scheme, and therefore no means of tracking the submission, all users are asked to supply an email address. If a form is submitted without contact details the message will be flagged by the system and the editor is provided with a warning. These terms may still be actioned at the discretion of the editor.

For further information see: [http://www.english-heritage.org.uk/thesaurus/thes\\_splash.htm](http://www.english-heritage.org.uk/thesaurus/thes_splash.htm)

## Appendix 4: Candidate term flowchart

