Guideline No. 15

Guidelines for Developing and Implementing a Keyword Thesaurus

February 2003

Summary

These guidelines cover the processes involved in compiling, implementing, reviewing and maintaining a thesaurus used for records classification purposes.

In these guidelines

1 Introduction

Purpose of these guidelines

The purpose of these guidelines is to assist staff of public offices and consultants to develop, implement, maintain and review a keyword thesaurus for use in records management.

The guidelines also provide information on the key concepts underlying the use of, and component parts of, a keyword thesaurus.

These guidelines form part of the framework of rules and guidance issued by State Records to help public offices meet their obligations under the State Records Act 1998. In particular, it aims to help each public office to meet their obligation to:

'…..make and keep full and accurate records of the activities of an office' (s. 12.1).

Use of a thesaurus is not mandatory.

Who should read these guidelines

The advice in these guidelines is relevant for public offices in all parts of the NSW public sector, including local government, the public health sector and universities. While these parts have thesaurus or classification scheme coverage through general retention and disposal authorities, the advice in these guidelines can assist public offices to customise and implement this. These guidelines will be useful for records managers, consultants and others involved in developing, customising, implementing or reviewing a keyword thesaurus.

Definitions

Key terms used in these guidelines are defined as follows:

Thesaurus: a controlled list of terms linked together by hierarchical, associative or equivalence relationships. (AS ISO 15489.2, 4.2.3.2).

Keyword thesaurus: a records management thesaurus based on functions and following the principles of keyword classification.

Keyword classification: involves grouping records into broad, functionally based areas represented by keywords. Records are further classified by the use of activity descriptors and optional subject descriptors.

Classification: systematic identification and arrangement of business activities and/or records into categories according to logically structured conventions, methods, and procedural rules represented in a classification system. (AS ISO 15489.1, 3.5).

For other, general recordkeeping terms, see State Records' Glossary of Recordkeeping Terms.

Acknowledgements

Parts of these guidelines have drawn on the National Archives of Australia's *Developing a Functions Thesaurus: Guidelines for Commonwealth Agencies*.

Contacts

For more information, please contact State Records.

2   Understanding Thesauri

Purpose of a thesaurus in records management

A thesaurus is a tool that supports the classification and management of records, usually at the file level. It ensures that classification terms are used consistently throughout a recordkeeping system. It is a 'controlled language' tool.

Why base a thesaurus on function?

Thesauri for records management purposes, which are based on business functions, enable records to be classified according to the context in which they are created and used. In contrast to a subject approach, records are therefore classified according to why they exist rather than what they are about.

Using a functions based approach for classification recognises that records are defined by their relationship to the activities they document. This gives records meaning and context.

Following this approach, the thesaurus is usually structured so that the first two levels in the thesaurus relate to functions and supporting activities. The third level may be free text or controlled subject descriptors.

Features of a thesaurus

A thesaurus is a list of controlled terms that is structured though relationships between terms. These relationships are explained in more detail in the following section, Concepts in Thesaurus Construction.

As a tool to title records, a thesaurus has a number of features that make it more user-friendly than its close relative, the Business Classification Scheme. A thesaurus may have:

- multiple entry points to guide users to preferred terms and correct titles
- cross-referencing
- scope notes and tips
- strict control of language, and
- alphabetical or hierarchical presentation.

When compiling a thesaurus, it is important to use the features that suit the particular implementation needs of your organisation.

Benefits of classification

Classifying records according to function brings many benefits for the management and effective retrieval of records.

The functional classification of records provides a structure to help determine and implement retention, security and access decisions, including sentencing on creation.

Consistent titling of records and the use of a consistent language throughout the organisation assist in records retrieval and the sharing of information across the whole organisation.

Functional classification establishes and documents the relationships between records and the business activities they document which is essential in understanding records, and in particular understanding over time.

Relationship with metadata

The use of a thesaurus or other controlled language tool can assist organisations to produce quality metadata and comply with the [New South Wales Recordkeeping Metadata Standard](). Thesaurus terms are generally used to title records. These terms will therefore populate the 'title' metadata field. Function elements also form part of the metadata standard and the thesaurus can assist with populating these fields.

Relationship with business classification schemes

A business classification scheme (BCS) is a representation of the functions, activities and transactions of an organisation. Creation of this scheme is outlined in Step B of the DIRKS methodology. It provides a conceptual map of the work your organisation performs. It can be a hierarchical or relational representation. Unlike a thesaurus, it is not used to classify or title records but does form the basis of a thesaurus.

Though a BCS and thesaurus are both built on functions, activities and transactions and share many common terms, they serve different purposes. A BCS is a conceptual map while a thesaurus is a records classification tool. They may be presented in different ways to serve these different purposes.

3   Concepts in Thesaurus Construction

Overview

This section explains key concepts and relationships in thesaurus construction. They are based on the international standard ISO 2788—1986 *Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri*.

Hierarchical relationships

A hierarchy is formed when a preferred term represents a concept which can be linked to another term with a broader or narrower meaning.

Broader term or BT indicates that there is a term with a wider meaning than the term given. Conversely, narrower term or NT indicates that there is a more specific concept than the one listed. The reciprocal relationship is shown below:

Maintenance   **BT** (or **Broader Term**) Property Management

Property Management   **NT** (or **Narrower Term**) Maintenance

In a keyword thesaurus for records management, the broadest terms represent functions of an organisation. They are referred to as keywords, for example, Property Management.

Activities that are carried out as part of that function are presented as narrower terms of the keyword. They are called activity descriptors. For example, maintenance is an activity supporting the Property Management function. Subject descriptors are further refinements of the activity descriptors, presenting aspects or topics of the activity. Using the Property Management – Maintenance example, the subject descriptor may be the name of the building being maintained, or difference aspects of the process, such as pest control or cleaning.

Equivalence relationships

The use of preferred and non-preferred terms is a characteristic of a controlled language thesaurus. Preferred terms are permitted terms that can be used to represent a given concept. Non-preferred terms, also known as 'forbidden' terms, are not used to classify documents and records. They are included in the thesaurus to act as pointers to preferred terms (ISO 2788 3.6).

This relationship between preferred and non-preferred terms is the equivalence relationship. When two or more terms can be used to refer to a given concept, one is selected as the preferred term to be used in the classification scheme (ISO 2788 8.2.1). Non-preferred terms are included in the thesaurus, as access points for users, as shown in the example below where 'additions' is a non-preferred term:

> Additions
>   **Use** Extensions

Relationships can also be created from preferred terms to non-preferred terms as shown below, using either the Use for convention or marking it as a non-preferred term (NPT):

> Extensions
>   **Use for** Additions
>
> OR
>
> Extensions
>   **NPT** Additions

As most users are concerned with finding the correct term, this reciprocal part of the relationship may not need to be shown.

Associative relationships

Association is used when preferred terms at the same level of the hierarchy are related in meaning and it would be useful to draw users' attentions to a similar term that may more accurately represent the concept they wish to describe.

Associative relationships are indicated by the 'related term' or RT direction, as shown below:

> Security

RT Privacy

The reciprocal relationship is also shown:

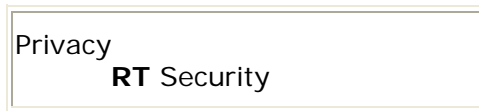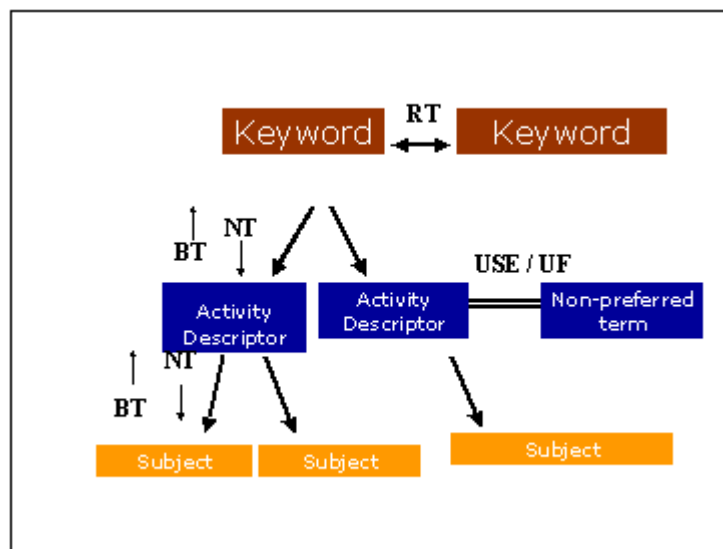| Privacy |
|---|
| **RT** Security |

Diagram of relationships

The diagram below shows relationships and common levels in a keyword thesaurus:

Scope notes

Scope notes are used to clarify the exact meaning of a term and also when to use it in the context of a thesaurus. It specifies what a term covers and excludes other possible meanings (ISO 2788 6.6.1).

SN can be used as a prefix to designate a scope note.

See references may be included in a scope note, usually following the main definition. Like related term links, they point the user to other classifications to consider. If the scope note for the term excludes one meaning, a see reference can point to the correct classification.

Dates may also be included in a scope note to show when it was first used in the thesaurus or when it stopped being a preferred term.

The **source** of the term may also be included to show whether it is from a published thesaurus such as *Keyword AAA* or the organisation's own functional research.

Date and source information should only be visible to thesaurus administrators rather than all users.

History notes

Date and source information may be included in a separate history note rather than in the scope note. This may also be used to document changes to the function or activity over time. Generally, history notes are not visible to users.

Tips

A thesaurus usually includes tips designed to help users select the right term or to title a record properly. As titles comprise both authorised terms from the thesaurus and free text, tips can provide advice on what to include in the free text. For example, the term 'Licensing' may be accompanied by the tip 'Include name of organisation being licensed as free text'.

Abbreviations and acronyms

Most organisations use a large number of abbreviations and acronyms. A thesaurus should specify what abbreviations and acronyms are permitted. This can be done through an appendix specifying permitted abbreviations and acronyms, and/or including them as preferred terms. Non-permitted abbreviations and acronyms can be included as non-preferred terms, with a reference to the full term. Abbreviations and acronyms, and their full versions, usually refer to concepts and entities that fit at the subject descriptor level.

The test as to whether an abbreviation or acronym should be permitted is whether it will carry meaning over time. For example, QANTAS will, whereas the acronym for a current project or current section name will simply cause confusion in the future.

Where an abbreviation or acronym is not permitted, it can be included in brackets at the end of the full term to facilitate searching, for example "Independent Commission Against Corruption (ICAC)".

4   Compiling a Keyword Thesaurus

Overview

Compilation Process

The process of compiling a keyword thesaurus can be broken down into six stages. These are:

Stage	Description
1	Preparation
2	Collecting information
3	Analysis
4	Collation
5	Seeking feedback
6	Production

Relationship to DIRKS

The process of compiling a keyword thesaurus has strong links to the DIRKS process. Stage 1, Preparation, may be carried out before commencing a DIRKS project, to help determine the necessary scope of the DIRKS project, or once Steps A, B and C of DIRKS have been completed. Stage 2, collecting information, refers to the information gathered in Steps B and C of DIRKS. These links are explained in more detail in each stage.

Other steps in the DIRKS methodology can help develop means by which your thesaurus can be incorporated into organisational recordkeeping systems and practices.

Use of software

The process described assumes that thesaurus compilation software is used, either a specific thesaurus tool or the thesaurus component of records management software. See Appendix B for details on how to select suitable software.

It is possible to compile a thesaurus manually or in a word processing program but this is very difficult and time consuming. This approach is not recommended.

Stage 1: Preparation

Identify need

The first task in a thesaurus development project is to establish a need for such a tool. While all organisations should have a classification scheme for their records, not all will require a thesaurus.

Your organisation may not need a thesaurus if it:

- is small in size
- controls records creation centrally,
- creates a small volume of records, or
- creates a large but narrow range of records.

In these situations, a simple authority list of file titles or file titling instructions may be sufficient.

Management support

Management support is important to the success of the project as implementation of a thesaurus will impact across the entire organisation, or all divisions that use a particular recordkeeping system.

Management support is also essential to provide the appropriate resources to carry out the project.

Planning and resources

Planning: The thesaurus project may be conducted as part of a broader DIRKS project, and therefore covered by DIRKS project planning. If it is being conducted

as a separate project, it must still be properly scoped and planned. In either case, required tasks and resources must be clearly identified.

Resourcing: Resources required for the project are staff with the appropriate skills and software for the compilation process. The appendices provide more detail on necessary skills and selecting appropriate software.

Alternatively, if consultants are being employed to undertake the project, appropriate funding, and the time of internal staff to liaise with the consultant and supervise the project will be required.

Stage 2: Collecting Information

Link to DIRKS

The primary information used to develop a thesaurus is collected though the DIRKS methodology, in particular steps A and B. In order to compile a thesaurus, a good understanding of the functions and activities of the organisation is required. This is gained through undertaking Step B of DIRKS.

Step C

Step C can also provide useful information for thesaurus construction. Step C involves identifying the recordkeeping requirements of your organisation. As a thesaurus is a file titling and classification tool, it is useful to have an understanding of requirements to create records and how records should be kept.

For example, in Step C you may identify a need to keep a number of activities under a particular function together as a case file. Therefore you would use the term 'cases' or another descriptor in your thesaurus rather than the activity terms as they appear in the business classification scheme.

While it is possible to compile a thesaurus after Step B, completion of Step C is strongly recommended.

Interviewing and consultation

It is important to consult with users at an early stage to help identify what information needs to be included in scope notes, whether terms are understood in the organisation and what tips should be included.

Stage 3: Analysis

Summary of stage 3

This stage involves analysing all the information collected in earlier stages to make key decisions about the features of your thesaurus. The table below outlines the key areas to consider:

Scope of thesaurus
Examining the BCS

Scope of thesaurus

The use of your thesaurus affects what must be included in it. Examine what your thesaurus will be used for. Possibilities include:

- file titling in a single recordkeeping system
- titling across a number of recordkeeping systems
- classification in other business systems, and
- providing structure for intranet sites.

Keep the intended use of you thesaurus in mind as you analyse information and develop the thesaurus.

Examining the BCS

It is necessary to re-examine and refine the business classification scheme to develop a thesaurus. The table below outlines required steps.

Step	Action
1	Decide what functions from the BCS should be included in the thesaurus *For example, not all functions may be documented in the recordkeeping system the thesaurus will be used in*
2	Ensure activities are useful, adequate and reflect how records should be kept *For example, some activities may be grouped together to form case files*
3	Ensure function terms (keywords) and activities (activity descriptors) are clear and understood by the intended audience
4	Using BCS notes, develop scope notes for thesaurus terms
5	Develop additional levels in the thesaurus *See below for more information*

Levels in the thesaurus

Two levels of hierarchy, keyword and activity descriptor, are a minimum. Many thesauri go down to a third, or more, levels. While your BCS will include a third, transactional level, in file titling, topic or subject descriptors may be more relevant.

Some parts of the thesaurus may require greater detail, and more levels in the hierarchy, than others. For example, subject descriptors may be useful under one function / activity pair, but instructions on the use of free text for another.

For example, one file title may be:

Keyword	Activity descriptor	Subject descriptor	Tip
Keyword Products	Licensing	Keyword for Councils	[name of council]

Another title, from the same thesaurus, may only have 3 levels:

Keyword	Activity Descriptor	Subject Descriptor
Training Services	Planning	Records Management Fundamentals

In this second example, the name of a course is used at the third level. However, if course names changes frequently, the use of a tip such as [name of course] may be easier to maintain than including all the names of courses as subject descriptors in the thesaurus.

Another consideration when deciding what levels to include is the limitations of particular software. In many applications, third level terms can be attached to any activity descriptor. However, in functional classification the third level depends not just on the activity but also the keyword: that is, third level terms should be attached to a keyword / activity descriptor pair, not just the activity descriptor. A hierarchical presentation of the thesaurus does not face the same problem, as terms are always displayed in their hierarchical context.

Tips can also be used to provide more detail instead of adding an additional level of fixed descriptors. For example, the tip 'Use name of property' after a range of activities (Acquisition, Maintenance etc) will provide consistency across a number of related property management files.

Features of the thesaurus

A thesaurus should be tailored to the environment in which it is to be used. The features described in Chapter 2 may or may not be used depending on organisational need and context. The table below outlines common environments and some thesaurus features that can be deployed to suit them.

Environment	Thesaurus Features
Decentralised file creation with minimal quality control	Very precise scope notes Many non-preferred terms Strong guidance on subject descriptors or lots of tips.
Sole responsibility for retrieval lies with user	Very precise scope notes Many non-preferred terms
Classification in business systems	Scope notes and tips note required as terms built into system
Used in Intranet sites	Both hierarchical view for structure and alphabetical view for browsing

	are useful.

It is also very important to consider the functionality of your proposed production format (see stage 6) when identifying required features. For example, if you intend on using the thesaurus only in records management software and do not produce any other version, and the records management software does not support related terms, there is no need to build in these features.

Stage 4: Collation

Procedures for collation

The table below outlines recommended procedures for collating the information to create a thesaurus. These procedures assume that *Keyword AAA* is being used to describe the administrative work of your organisation and that software is being used to develop the thesaurus. If *Keyword AAA* is not being used, start from step 2.

Step	Action
1	Load *Keyword AAA* into software
2	Enter functional keywords as preferred terms. Enter scope notes
3	Add additional activity descriptors (that were not loaded with *Keyword AAA*) and scope notes
4	Create broader term / narrower term links between keywords and activity descriptors
5	Enter third and additional levels
6	Create links to additional levels
7	Enter non-preferred terms and link to preferred terms
8	Create links to related terms *Note that related terms must both be preferred terms and at the same level in the hierarchy*

This procedure does not have to be strictly followed. You may prefer to enter all the terms and scope notes associated with one keyword, and then move on to the next. However, to ensure consistency with *Keyword AAA* if you are using it, it is strongly recommended that you load it into your software before adding other terms. *Keyword AAA* is available in formats compatible with a number of thesaurus applications and records management software packages.

Using *Keyword AAA*

Keyword AAA is a thesaurus which covers the general administrative functions of government. Your business classification scheme and thesaurus will probably include terms from *Keyword AAA,* particularly some of the *Keyword AAA* activities in combination with your core business functions. You may also find terms in *Keyword AAA* that you wish to alter or change to suit your organisation's practices.

However, as *Keyword AAA* terms provide the structure for general retention and disposal authorities (GDAs) issued by State Records, altering *Keyword AAA* terms may require you to map disposal coverage in the GDAs to ensure it fits with your use of the terms.

The table below outlines issues to be aware of when altering *Keyword AAA:*

If....	Then...
You amend *Keyword AAA* scope notes to include organisation-specific meanings	Your organisation's use of the term will be different to that used in the GDAs. Some records titled with that term may be covered by GDAs and others will not.
You use keywords in *Keyword AAA* to describe you organisation's core business functions	GDA coverage of that term will not apply. You will need to seek additional disposal coverage.
You use activity descriptors in *Keyword AAA* to describe higher-level functions and change the hierarchy in *Keyword AAA*	You will have to carefully map disposal coverage and consider how you will title administrative records.
You create a new activity and link it to a *Keyword AAA* keyword	It will not be covered by a GDA. You will have to seek additional disposal coverage.

Selecting non-preferred terms

If necessary in your thesaurus implementation, non-preferred terms can be selected to help users navigate to preferred terms. Non-preferred terms can be identified by:

- checking search engine logs, i.e. checking terms used for searching
- examining organisation-wide publications such as strategic plans and annual reports for terms in common use
- examining previous file titling practices
- discussing with staff terminology which may be used in business units which may not be reflected in official sources such as annual reports
- assessing terms that were rejected as descriptors for functions and activities, and
- looking up terms in a standard thesaurus and selecting those likely to be misused.

Stage 5: Seeking Feedback

Why seek feedback?

Like any product, it is important to ensure that users will accept the thesaurus. Feedback from users will help highlight required changes, areas of confusions and any inconsistencies. It will also help you judge how much training is required to effectively implement the thesaurus.

Who to consult with?

The thesaurus should be circulated to the intended audience: is it a tool to be used by end users, or just records management staff? Are administrative assistants in business areas primarily responsible for filing? Use such questions to determine the range of staff from whom you would like to obtain feedback.

How to consult

As a thesaurus can be a complicated tool, it is useful to hold focus groups and meetings rather than just forwarding to staff for comment. Focus groups will give you an opportunity to explain how the thesaurus is to be used and ask questions on particular aspects. Areas to question may be:

- Is it clear how the thesaurus is to be used?
- Do the scope notes make sense?
- Can you see any gaps in coverage?
- Can you see any overlaps?

It may also be beneficial to give users a trial file titling exercise to help reveal any problems.

Focus groups and user meetings also give you an opportunity to sell the benefits of the thesaurus and answer any questions users may have.

Stage 6: Production

Production options

A thesaurus can be produced in many ways and you may choose to have more than one format. Options include:

- using in records management software
- using thesaurus browser application
- producing hard copies
- producing extracts for specific business areas in hardcopy or electronic format
- integrating into work processes, and
- loading onto intranet.

Production and implementation

Producing the thesaurus is both the final stage of developing it and the first stage of implementing it. The produced thesaurus must be seen as part of an implementation package that also involves training and user guidance.

Two issues to consider when producing your thesaurus are:

- the functionality of the proposed format, and
- the audience of the thesaurus.

The format of the thesaurus must match the needs of users and a number of different production options may be used. Different business units in particular may have different needs and guidance should be tailored for these groups.

For example, one business unit may carry out a variety of different functions in fairly unstructured, project-based work. They will probably need access to the entire thesaurus, plus advice on a common set of files for each project – a planning file, a consultation file, etc. Another business unit may do more process-based work that relates to a single function. A list of standard file titles, or building file creation and titling into the standard work process, may suit this unit.

Records management software

It is useful to load the thesaurus into records management software as its use can then be made mandatory. However, some records management software may not permit the inclusion of non-preferred terms or 'related' relationships that limit the functionality if the thesaurus. In addition, search functionality may also be limited. Therefore you may wish to support this with other versions in different formats.

Records management software usually does not allow more than one version of a thesaurus to be loaded at a time in the database and so cannot be used for documenting changes to the thesaurus over time.

Thesaurus browser application

Thesaurus compiling software can also be used to display the thesaurus to users. It usually has good search functionality and can of course use non-preferred and related terms to advantage.

If this approach is adopted, the software must have access controls so only authorised staff can make changes to the thesaurus.

A disadvantage of this approach is that multiple licences will have to be purchased to run the software on every desktop.

Another disadvantage is that, if users are responsible for creating files, the browser must be linked to the file creation application as users are unlikely to open and use two programs. However, if users then request a file to be created by a different staff member, then using the browser application, regardless of whether it links to the records management system, may be appropriate.

Hard copies

Many users prefer browsing through a hard copy publication. An alphabetical layout is very intuitive to use and allows the advantages of non-preferred terms and related links. A hierarchical presentation can be used to show terms in context.

However, hard copies are more difficult to keep updated. Publication dates and version numbers should be included.

Extracts for business areas

As people in particular business areas usually perform work relating to a small number of functions and activities, it may not be necessary for them to work with the entire thesaurus.

There are a number of ways to prepare extracts for particular business areas. You may:

- link thesaurus display in the records management software to user log-on so users are only presented with relevant portions
- create a 'cheat sheet' of common file titles, or
- present a limited form of the thesaurus in either hard copy or thesaurus browser.

It must be remembered that users will occasionally need to use other terms in the thesaurus, particularly administrative ones relating to such things as financial management or strategic planning.

Integrate into work processes

The classification of records can be incorporated into work processes so users are required to make fewer decisions. This may be done by incorporating records creation and classification into automated workflow systems.

Procedure manuals for various work processes can also be amended to include steps stating when to create a record and how to title it.

Display on intranet

The web functionality of an intranet offers many possibilities for displaying the thesaurus in a form that can be easily browsed or searched, and is easier to keep updated than hard copies.

5   Implementing a Keyword Thesaurus

Introduction

Links with other stages

Implementation issues should be considered throughout the development process. In particular, implementation and production are closely related. How you intend implementing your thesaurus will affect the choice of production formats. In addition, consultation during the development of the thesaurus, and promotion and communication throughout the project, will contribute to the success of the final implementation.

DIRKS Steps F and G

Step F and G of the DIRKS manual covers general implementation issues and processes such as planning, training, communication and change management. To avoid duplication, this section focuses on issues that relate specifically to implementing thesauri, in particular legacy file titling systems. Use DIRKS Steps F and G for advice on the following implementation tasks:

- allocating responsibilities
- allocating resources
- identifying other areas for improvement
- developing an education strategy
- preparing the implementation plan
- implementing, and
- training users.

Implementation strategies

Introduction

When implementing a new thesaurus, decisions must be made regarding how to move from a previous file titling system to the new one. Some conversion strategies with their advantages and disadvantages are outlined below.

Closure and recommencement

In this strategy, the records system is closed off at a particular date and a completely new system is started. This approach has a number of advantages:

- it creates less work for staff than trying to reclassify files (see next option)
- if there have been poor control mechanisms (i.e. classification and numbering) it is easier to close down the system and start again, and
- it avoids confusion amongst users and records staff as you actually close down the whole system and move to a new system.

However, it also has a number of disadvantages:

- many new files will need to be created and registered in a short period
- rigorous monitoring of files is required to ensure that no further documents are added to a file once the old system has been closed
- there is a lack of continuity between records systems, as the only link between the two systems is one of intellectual control, and
- it may lead to the creation of duplicate files. For example, if you had a 1998/9 Budget file and closed down the system on 1 January 1999 - you would have two files to cover that financial year.

Retrospective conversion

Retrospective conversion is when you reclassify old files using the new keyword thesaurus. We do not recommend this strategy. Its main advantage is that it provides the continuity that is not apparent with the closure and recommencement of a record system.

In contrast, its many disadvantages include:

- it is labour intensive. You need to decide how retrospective the conversion will be (for example, just files in the storage area, or only the last 2 years of files)
- existing files may not 'fit' the new scheme, for example, when more than one function or activity is documented in a single file
- links previously established between files may be destroyed, and

- disrupting the original order, which for future users gives context and aids in understanding the records.

Parallel systems

This involves creating new files as they are needed but not closing down the old system; you just gradually retire the older files. This means that there will be two systems running concurrently for a period.

This approach is suitable if you are not making any other changes to the control mechanism of the records system. It does not involve as much up-front work as other options.

Its disadvantages include:

- having two systems running concurrently may cause confusion among users, and
- it will take users longer to adapt to the new approach.

Piloting the thesaurus

Trialling the thesaurus with a pilot group before implementing to all users can be very beneficial. It gives an opportunity to fix problems before the whole organisation is exposed to it. If you start with a section of the organisation that is more likely to be receptive and enthusiastic to the thesaurus, they will become ambassadors - spreading the message regarding the benefits of the thesaurus to the rest of the organisation.

Maintaining links between systems

Regardless of the particular implementation strategy chosen, it is vital to maintain links between the old and the new systems. This includes creating links between old and new files that document the same matter, and documenting all changes to the records system.

The next section on reviewing and maintaining a thesaurus provides more information on what to document.

Implementation problems

Common issues

The table below outlines a number of common implementation problems and ways to address them:

Problem	Solutions
Users confused by large number of terms in thesaurus	Provide lists of common file titles Restrict display to terms they are likely to use
Insufficient records staff to answer day to day problems and queries	Train super-users in each business section to provide front line support

Remote or decentralised offices creating inconsistencies in classification	Provide lists of common file titles
	Build file titling into work processes
	Bring staff together for training to share ideas
Staff reluctant to use thesaurus	Promote benefits
	Provide enough training
	Use standard file lists and cheat sheets to make it as easy as possible

6   Reviewing and Maintaining a Keyword Thesaurus

Overview

Importance of ongoing management

A thesaurus is a living document that will require ongoing maintenance and periodic reviews to maintain its relevance and usefulness to the organisation. As organisations constantly change, so does the work they do and the language used to describe it.

Review or maintain?

A comprehensive review will only occur at certain intervals – perhaps every two years. Decisions must be made about what changes can be made on an ad-hoc basis, as part of on-going maintenance, and what changes should wait for, or trigger, a thorough review.

The scope of changes will depend on the expertise of the officer responsible for the thesaurus. If the officer is very familiar with the structure and content of the thesaurus and the work of the organisation, more changes may be made as part of ongoing maintenance. The table below provides some guidance on the split between maintenance and review activities.

Maintenance	Review
Adding non-preferred terms	Adding keywords
Adding additional tips	Amending existing keyword or activity descriptors
Adding activity descriptors or any lower-level term	Deleting any terms
Changing scope notes to clarify the meaning of a term	Changing scope notes to alter the meaning of a term
Adding related term links	Changing a preferred term to a non-preferred term

	Changing a non-preferred term to a preferred term

Business classification scheme and disposal authorities

As a thesaurus is based on your organisation's business classification scheme, changes to the thesaurus whether through maintenance or review should also be considered in conjunction with the BCS. This is particularly important when reviews are triggered by changes in the functional responsibility of your organisation.

Changes to the thesaurus may also break links to retention and disposal authorities. When making changes to the thesaurus, note where you may need to seek new disposal authorisation for added functions and activities.

Managing maintenance

Guidelines

Guidelines should be prepared that outline correct procedures to follow when maintaining the thesaurus. They should specify:

- what changes can be made (and what must wait for a review)
- who can approve changes, and
- what changes should be documented.

Requesting changes

Requests for changes often come from staff when the thesaurus does not meet their needs. The request for change should be accompanied by a brief explanation of why the change is required.

When a staff member requests a change, it is important to be sure that the change is appropriate and will not adversely affect the classification of other records.

Approvals

Responsibility for approving changes must be clearly identified. It may rest with the records manager, or a thesaurus advisory committee could be established to oversee ongoing maintenance. If super-users were identified and trained as part of implementation, they will form a good resource for either suggesting or approving changes.

Documenting changes

Any changes made to the thesaurus must be clearly documented. Notes about requested changes that are not implemented should also be kept for consideration in a future review.

Managing a review

When to review?

A review may be required for any of the following reasons:

- the organisation loses or gains a function
- business culture changes and new terms come into use
- State Records releases a new edition of *Keyword AAA*, or
- users identify a number of weaknesses that indicate an overhaul rather than minor changes are required.

Even when there have been no obvious triggers for a review, it is still good practice to review the thesaurus on a regular basis.

What to review?

The following factors should be considered when reviewing the thesaurus:

- terms are relevant and reflect the work of the organisation
- descriptors are not duplicated, scopes do not overlap and relationships with other terms are still relevant
- the frequency with which descriptors are used (most records management software can do this automatically). If descriptors are under- or over-used, investigate why, and
- abbreviations and acronyms are referenced correctly and used properly.

Usually it is preferable if someone who was not closely involved in the development process conducts the review. As with development, user consultation is essential to the success of the review.

Documenting the review

As well as using existing documentation to conduct the review, such as staff suggestions, it is also important to document the review to justify changes and provide information for future developments.

Both the original and new versions of the thesaurus must be retained, as well as explanations for changes. Also document any changes made to *Keyword AAA* terms. If you do not document these changes, you may have difficulty remembering what was done and identifying changes needed to new versions of *Keyword AAA*.

Implementing changes

Conversion strategies

The conversion strategies outlined in Implementation also need to be considered. Depending on the scale of the review, implementing changes to a thesaurus can be almost as large a task as implementing a new one. A parallel conversion or closure and re-commencement may be the most suitable option. However, if the changes are minor, it may be viable to re-classify the affected records.

Software issues

Care must be taken when implementing the new thesaurus. As most records management software can only hold one version of the thesaurus, make sure to

take a printout of the old thesaurus for record purposes before making changes. If you are using thesaurus software, it should be able to store multiple versions.

Training staff

As with the initial implementation, staff should be informed of changes to the thesaurus. Depending on the extent of the changes, and the amount of staff turnover since training was last offered, formal training may be required.

Updating documentation

The thesaurus itself, and any documentation associated with it such as common file title lists, extracts or work procedures must be located and updated to prevent confusion.

Appendix A: Skills of the Thesaurus Compiler

Options

Compiling a thesaurus is a specialised task in records management. Organisations have two options:

- acquiring expertise by employing someone with the necessary skills or having someone trained, or
- using consultants.

Acquiring expertise

This option has an advantage in that the organisation can use the expertise over time for reviewing and updating the thesaurus, training staff in its use and quality checking. By training an existing officer, it is possible to build on their knowledge of the organisation's functions. This approach also supports the professional development of staff and encourages ongoing and informed maintenance of the thesaurus.

Using consultants

This approach has the advantage or high level specialist skills and experience. However, consultants need to become familiar with the organisation. Also, the deep understanding the compiler has of the thesaurus will be lost at the conclusion of the project.

Skills

The skills required to compile and implement a thesaurus include:

- knowledge of the organisation and its functions, or the ability to quickly acquire this

- written and oral communication skills

- analytical skills

- ability to liase with stakeholders

- knowledge and understanding of keyword classification and thesauri rules

- knowledge and understanding of the functional approach in recordkeeping

- knowledge of project management methodologies

- ability to promote and market the use of a thesaurus

- ability to train users

- ability to educate other staff about thesaurus compilation and maintenance, and

- abilities to work with little supervision and to use initiative.

All these skills do not need to be found in the one person. For example, one person may be responsible for managing the project, one for training and implementing and another for actually compiling the thesaurus.

It is important to ensure that the compiler is given dedicated time and resources to the project. Compiling a thesaurus requires intense concentration, so it is not suitable to try and complete it in spare time snatched from other work.

Appendix B: Selecting Thesaurus Software

Purpose

The purpose of this appendix is to outline features of thesaurus software. The importance of each category of features will depend on whether you intend using the software for compiling a thesaurus, managing the thesaurus and/or as a search interface for all staff to use.

Validation and data entry

Examples of requirements relating to validation and data entry in a thesaurus application include:
- a term cannot be linked to itself
- can handle sufficient number of levels in a thesaurus
- a non-preferred term cannot have broader, narrower or related term links
- a relationship needs only to be entered once and the software will create the reciprocal relationship (BT/NT, RT/RT)
- fields are included to indicate source of term, and
- scope notes and history notes can be kept separately.

Data management

Requirements may include:

- display tags can be changed so each level can be given an appropriate title (for example, Keyword, Activity, Subject or Function, Activity, Transaction)
- application includes access control to ensure only authorised staff can change the thesaurus
- application can maintain different versions of the thesaurus, or can be able to reconstruct a thesaurus at a particular point in time, and
- application can export data in a variety of formats so that it can be imported to records management software or published on an intranet.

The application should have a variety of display and report options for online and hardcopy. It is also desirable to enable ad hoc reports, which the thesaurus manager can easily produce. These might include:

- full thesaurus in alphabetical order
- authorised terms in alphabetical order, with or without scope notes
- authorised terms in hierarchical order, with or without scope notes
- terms from a particular source, and
- portion of the full thesaurus.

Search and retrieval

Depending on how the thesaurus will be implemented, the search and retrieval functionality of the thesaurus software may be important. Requirements may include:

- more than one access point to the thesaurus. May include hierarchical and alphabetical browsing, and a search function.
- ability to search on term without knowing for sure if it is in the thesaurus, and
- ability to integrate with other applications. For example, if users select the terms from the thesaurus software can it populate the appropriate data element in the recordkeeping system?

Further Reading

Australian Standard AS ISO 15489 - 2002, *Records Management.*

Australian Standard AS 4390 - 1996, *Records Management.*

International Standard ISO 2788-1986 *Documentation – Guidelines for the Establishment and Development of Monolingual Thesauri.*

National Archives of Australia, *Developing a Functions Thesaurus: Guidelines for Commonwealth Agencies.* Available online at
http://www.naa.gov.au/recordkeeping/control/functions_thesaur/intro.html

New South Wales Office of Information Technology, *Project Management Guidelines*. Available online: http://www.oit.nsw.gov.au/pages/4.3.21.Project-Management.htm

Robinson, Catherine and Knight, Janet, 'Contemporary Recordkeeping: The records management thesaurus – A response' *InfoRMAA Quarterly* 14/1 February 1998.

State Records Authority of New South Wales, *Designing and Implementing Recordkeeping Systems (DIRKS) manual*. Available online: http://www.records.nsw.gov.au/publicsector/DIRKS/exposure_draft/title.htm.

State Records Authority of New South Wales (1998), *Keyword AAA: A thesaurus of general terms*, Sydney.