Government of South Australia

GPO Box 2343
Adelaide SA 5001
Tel (08) 8204 8773
Fax (08) 8204 8777 DX:467
srsaRecordsManagement@sa.gov.au
www.archives.sa.gov.au

State Records of South Australia

Developing a Thesaurus

Guideline

June 2002

Version **1.3**

Table of Contents

Introduction

A thesaurus is an alphabetical listing of terms derived from a classification scheme and arranged in a hierarchical manner that shows relationships between the terms. It prescribes what terms can and cannot be used when classifying records.

A thesaurus helps to ensure that an agency uses a uniform set of terms to classify records. This cuts down confusion when searching for records and saves time and money.

This guideline has been developed to assist agencies in complying with Outcome 5 of the *Adequate Records Management Meeting the Standard*– ‘Records can be found’.

Agencies are legislatively required to demonstrate adequate records management programs and practices in accordance with the *State Records Act, 1997*. Effective classification of records, through the development and use of a thesaurus, facilitates capture, retrieval, maintenance and disposal of records.

Scope of this guideline

This guideline has been developed by State Records to provide agency staff with an outline of the steps and issues inherent in developing a functional thesaurus.

This guideline applies to agencies within South Australia as defined in section 3 of the State Records Act which includes state government agencies, local government authorities and universities (refer to the glossary for definition).

This document is issued as a guideline in accordance with section 7(g) of the State Records Act.

Structure of this guideline

The guideline is structured accordingly:

- Part 1 – Developing a thesaurus
- Part 2 – Functional or subject based thesaurus?
- Part 3 – Integration of agency and general thesauri
- Part 4 – Linkages between thesauri and disposal schedules.

Variation to this guideline

State Records may update or alter this guideline from time to time as authorised by the Manager [Director] of State Records, in consultation with the State Records Council. All South Australian agencies will be informed of any such alterations or updates.

Acknowledgments

The following material from the State Records Authority of New South Wales has been used in developing this guideline:

- Compiling a functional thesaurus to merge with Keyword AAA

- Compiling a Keyword Thesaurus training course
- Keyword AAA – A thesaurus of general terms.
- Important note

Agency staff using this guideline should be familiar with the concepts and principles of using and developing a thesaurus. State Records offers two training courses in this area:

- Using a Keyword Thesaurus to Control and Classify Records
- Developing a Functional Thesaurus.

Many of the areas touched upon briefly in this guideline are explored in greater depth during the training courses.

This guideline does not replace the training courses.

Staff intending to undertake such courses should contact the Training Coordinator at State Records; www.archives.sa.gov.au/contact.html.

Further information

Agencies and authorities who seek information regarding any of the principles in this Guideline should contact the Manager, Government Recordkeeping at:

State Records of South Australia
GPO Box 2343
ADELAIDE SA 5001
Phone (08) 8204 8773
Fax: (08) 8204 8777

Thesaurus types

Essentially there are three types of thesauri:

- agency specific (or operational)
- general administrative
- merged.

The first type is developed by, or for, specific agencies and addresses records that are only created by that agency and are ‘operational’ or ‘business specific’ to the particular agency. This type of thesaurus can be developed along functional or subject-based lines to address the operational functions and activities of the agency as evidenced in its official records. State Records recommends agencies use functional based thesauri when developing a thesaurus specific to their own records (see Part 2).

The second type of thesaurus is generally developed by a central agency for application across an entire sector. An example of this is the *Keyword AAA Thesaurus of General Terms* developed by the State Records Authority of New South Wales and adopted for use in South Australia. Increasingly thesauri in this category are being developed using functional approaches.

The third type combines terms from the operational and the general administrative thesaurus into a single merged thesaurus of terms.

Part 1 – Developing a thesaurus

The process of developing a thesaurus can be divided into the following ten distinct stages:

- prepare
- collect information
- analyse
- collate information
- check terms against any existing sector wide thesauri
- compile draft
- circulate draft for comment
- produce completed thesaurus
- implement
- review.

Stage 1 – Prepare

Explanation

The development of a thesaurus is a complex and involved process. It requires dedication of time and resources by an agency to ensure a successful outcome. It is therefore critical that the project is well prepared by addressing the following points:

Establish the need

A number of key indicators can help to determine whether there is a genuine need or business case for a thesaurus in your agency:

- staff are unable to find records even though they know that they have been registered and indexed
- an agency's records management staff become aware that terms are used inconsistently
- particular sections of an agency are consistently reporting problems in finding records
- records of a particular area are unable to be accessed and found by other staff of the agency
- disposal schedules are difficult to apply to records because records with varying retention periods are placed on the same file.

At this point take the following steps:

- survey agency practices (including consultation with agency staff) to define the scope of the problems that initiated the project including:
  - the number of users classifying and indexing using variations on a common term
  - the number of instances where users are unable to find records, and the reasons why
  - acronyms and industry terms that are adding complexities to finding records
  - local user requirements for particular terms, such as acronyms or industry terms
  - inspect records and recordkeeping systems.

- identify problems caused by current record classification systems[1]
- document the findings.

Obtain support

Once you have established that there is a need for a thesaurus, obtain the relevant management level support for the project. Without this support the project will achieve only poor outcomes or will fail. Support your case by:

- briefing managers on your findings from previous point above
- developing a business case for the project
- acquiring project sponsors within the management structure.

Select team

The selection of a team to work on the development of a thesaurus will be driven by a number of factors including:

- the size of the project
- the human resources and skills available within your agency
- the degree of management support received.

Consider whether or not to use a consultant to either assist with or conduct the project.

Should your agency decide to use in-house staff for the project then back-filling their positions needs also to be addressed. When selecting a team consider:

- the skills of the team members
- the team’s knowledge of the agency and its functions
- their experience in using the agency’s records
- appropriate classification levels
- the commitment of the project sponsor
- available funding.

Develop project plan

The complexity of developing a thesaurus requires a project plan which includes:

- project scope and objectives
- criteria for quality measurement
- required outcomes
- timetables
- specific activities and tasks

[1] Key indicators of problems with any existing classification systems include:

- uncontrolled and inconsistent use of language in file titling and indexing
- inefficient and inaccurate searching and retrieval of information and records.

- responsibilities
- indicative costs
- a communication strategy.

The project plan needs to comply with any existing agency-specific templates or frameworks.

Stage 2 – Collect information

Explanation

The preparation of a thesaurus requires the project team to identify and collect a range of information that is central to establishing an understanding of the agency's functions and activities. The following steps should be followed by the project team:

Examine the documentation

The primary source of information about the agency and its functions should be agency documentation. State Records recommends identifying and analysing:

- annual reports
- strategic plans
- corporate business plans
- enabling and controlling legislation
- corporate website and/or intranet site
- standards
- codes of best practice.

Each of these sources will provide certain information useful to the project team including:

- annual reports – some detail of the major functions of the agency and their component activities, past achievements, planned future development and functions, and information regarding statutory or other establishing documentation
- strategic plans – an indication of the planned future direction of the agency and some summary information about the activities to be undertaken to achieve the stated goals
- corporate plans – objectives that the agency has identified as priorities (these can be used for potential keywords or initial scope notes)
- enabling and controlling legislation – core business processes and record creation requirements, and record retention and access requirements
- corporate website and/or intranet site – agency structure and division of tasks and responsibilities, and functions and activities of the agency.

Identify any existing disparate and sector-wide thesauri

The project team needs to identify existing relevant disparate and sector-wide thesauri; the most common one will be *Keyword AAA: A Thesaurus of General Terms.* Agencies that are members of technical or professional bodies are advised to check with such bodies to establish if any thesauri of common technical or professional terms already exist.

Check for any existing records disposal schedules

The project team should also establish if any records disposal schedules exist for the agency. If so, the team should establish if any existing schedules were developed along functional lines. If not, the team should investigate whether the existing schedules are due to be replaced in the near future, and, if so, the team would be well advised to create the functional thesaurus at the same time as developing a new functional records disposal schedule. The functional analysis and business analysis performed for one exercise will be of great benefit to the other.

It is recommended that a records disposal schedule be developed using the same function and activity terms as the thesaurus; this will assist in the records disposal process.

Interview selected managers and staff

As well as accessing documents about the agency the team should also make use of expertise on the business of the agency among the agency's staff. State Records recommends that the team should ensure that the staff it interviews are:

- in positions that cover a range of functions and activities
- aware of the project and its outcomes (this can be facilitated by the development of a communication strategy for the interviews).

Questions to be asked by the team should include the following:

- What does your area do?
- What do you do in the conduct of your business?
- How would you break these broad areas down?
- Do you work with other areas?
- What terms do you use to identify your functions?
- What terms do you use to identify your activities?
- What documentation do you create?
- What legislation do you operate under?
- What reporting requirements are you aware of?

Stage 3 – Analyse

Explanation

When the project team analyses the documentary sources and the information gathered during the interviews, they will be able to identify the functions, activities and transactions of the agency.

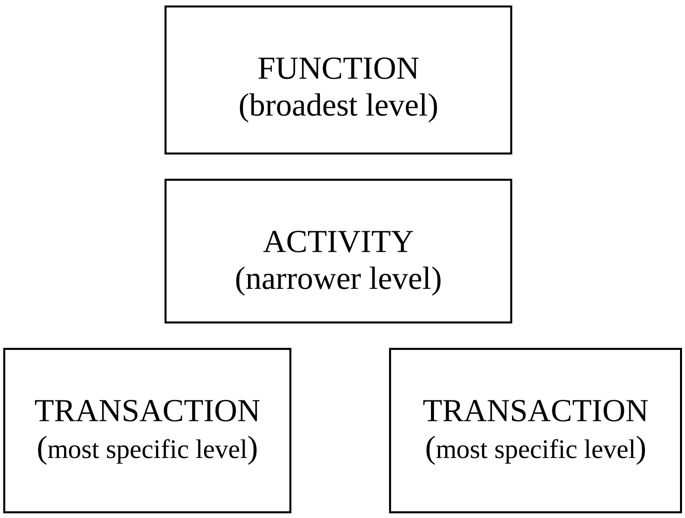
Analysis of business activity using documentary sources

The team should also use the documentation to ensure that it has identified the correct terminology for use in the thesaurus. This is especially important in areas where technical terms are in common use. It will also help to reduce confusion where agency terms are similar to those in sector-wide thesauri (see Stage 5 and Part 2 below).

Once the relevant terms, functions, activities and transactions are identified they are placed in a hierarchy to form the business classification scheme that should:

- include all of the business functions and activities being documented
- be linked to functions and activities, not organisational units
- be described using unambiguous terminology
- be hierarchical (moving from the broad {function} to the specific {transaction}
- be specific to the organisation
- be devised in consultation with users
- be maintained to reflect changing business needs.

The links between functions, activities and transactions can be seen in the following diagram[2]:



[2] An agency will have more than one function. A function can have more than one activity, and an activity can have multiple transactions. An activity can be linked to more than one function (not shown above because of space constraints).

Confirm draft business classification scheme

Since the business classification scheme will form the building block for the thesaurus it is essential that the team confirm the scheme with staff of the agency by forwarding the relevant extract of the draft business classification scheme to the staff initially interviewed. At this stage the project team should be seeking to confirm that:

- the scheme accurately portrays the business functions, activities and transactions of the business unit
- the draft scheme is comprehensive in its coverage of the business functions, activities and transactions of the business unit
- the business unit understands the terms used to describe its functions, activities and transactions.

Stage 4 – Collate information

Explanation

It is at this stage that the project team starts to build the actual thesaurus, working from the completed analysis in Stage 3 of the process. If your agency is using the *Keyword AAA* thesaurus it will make an efficient model on which to build your functional thesaurus.

It is vital that the project team ensures that they do not replicate functional keywords from the *Keyword AAA* thesaurus directly into its functional thesaurus. It is possible, however, to replicate activity descriptors, as an activity can be linked to multiple functions that are both government wide and agency specific.

Activities that are included in the functional thesaurus and already exist in *Keyword AAA* need to be used in the same context. The scope note gives the term its meaning and context. The business activity has one meaning but can be used within a number of functions.

Some functions identified in the agency may also be identified in *Keyword AAA* but used in a different context. In this case it is important that different keywords to those in *Keyword AAA* are used for the functional thesaurus where possible.

For example, ’Grant Funding’ is used as an activity descriptor in the *Keyword AAA* thesaurus. However, Grant Funding may be a core business function of an agency. In this case the agency needs to identify an appropriate term to represent the core business function and ensure that adequate ’tips’ and ’see’ references are included underneath both terms to ensure appropriate use.

The team should work through the following processes:

- identify preferred and forbidden terms[3]
- organise preferred terms into a logical hierarchy

[3] These terms are defined in Part 6 of the International Standard 2788. See also the glossary to this guideline.

- develop keywords, descriptors and their associated scope notes, ensuring their exclusiveness and including relationships and cross references between terms
- establish and verify links between terms.

Stage 5 – Check terms against other thesauri

Explanation

Once the project team has created and compiled the terms intended for use in the thesaurus, it is essential that the terms be checked against other thesauri being used in the agency. This will reduce confusion among agency staff when trying to find the correct term. This will also reduce confusion when the thesaurus is used in an automated environment. The check will identify:

- duplication of terms
- overlap of terms
- the need for the creation of alternate terms where necessary
- requirements for references where terms are similar.

In instances where terms are similar the project team should insert explanatory text or 'see' references to enable users to identify the intent behind the terms and their placement in the thesaurus, e.g.:

TRAINING SERVICES

see TRAINING for staff training.

Stage 6 – Compile draft

Explanation

Having established that there is no conflict, confusion or duplication between existing general thesauri and the terms to be used in the agency thesaurus, the project team should now compile a draft of the thesaurus. This can be done either manually or by using thesaurus compilation software. The draft will require printing out and proof reading. State Records suggests that staff other than the project team do this, where possible. This stage of the process will also enable the project team to pick up typographical errors and spelling mistakes.

Stage 7 – Circulate draft for comment

Explanation

Once a draft of the thesaurus has been developed it should be distributed for comment. This will ensure that staff within the agency are kept informed and will also give them a means for correcting mistakes or areas of confusion. It may be necessary to amend, revise or update the thesaurus as a result of the consultative process. During the development of the draft the team should also:

- provide each of the staff they interviewed with a copy of the relevant section of the draft for review and comments
- make any changes to the draft deemed necessary following the review process.

Stage 8 – Produce a completed thesaurus

Explanation

Once the consultative process has been completed the project team will need to produce the release version of the thesaurus with due consideration of:

- adherence to corporate documentation appearance standards
- adherence to corporate documentation approval standards
- the method of distribution, i.e. in paper form, in electronic form or via a network or corporate intranet.

Stage 9 – Implement

Explanation

This stage is about communicating the completion of the thesaurus and making plans for its implementation by establishing or developing:

- consultation mechanisms with users
- a conversion process to the new thesaurus
- a timetable for implementation
- a conversion strategy, where existing thesauri are being replaced
- a training program for users, targeted to meet their needs and use of the thesaurus
- feedback mechanisms
- help desk and/or other user support mechanisms
- a register or record of issues and problems as they are identified by users
- the rollout mechanism for the thesaurus i.e. pilot application, phased or global rollouts
- a program of testing the product and processes.

The agency will also need to develop and distribute:

- a maintenance plan
- policy statements
- user guides
- rollout procedures
- performance measures
- support systems.

The *Australian Standard Records Management AS ISO 15489.1-2002*, subclause 8.4 and *AS ISO 15489.2-2002*, clause 3 provide a methodology for designing and implementing a records system. This methodology might help to establish an implementation plan for the new thesaurus. However, the methodology in the standard was not developed for implementing a thesaurus, and the team should take care in using it to develop the thesaurus implementation plan.

Stage 10 – Review

Explanation

As with all complex projects a post-implementation review is essential. It enables the agency to identify problems that have arisen from the practical application of the thesaurus. The project team will need to establish a review date and a review team. State Records also recommends that the agency develop and instigate a thesaurus maintenance program.

Part 2 – Functional or subject-based thesaurus?

Explanation

When developing a thesaurus agencies should select the structure that best suits their needs. State Records preferred option is for thesauri developed along functional lines. However agencies should contact State Records – before the project starts – if they are of the opinion that a subject-based thesaurus is more appropriate to their particular circumstances.

Over the past few years there has been an increasing shift within the records management sphere from subject and class based thesauri to functional thesauri. This change has been driven by the view that while agency structures may change, the agency's essential functions remain the same. By structuring thesauri on functional lines, records managers have ensured the ongoing viability of their thesauri in a rapidly changing environment.

Another factor driving the change in approach has been the emergence of functional records disposal schedules, previously developed along subject or record class lines. This change has also been driven by the view that schedules structured in this manner will remain viable for longer. With the increasing integration of thesauri and records disposal schedules in automated records management software, there is a critical need for similarities in approach to the development of records disposal schedules and thesauri. Without integration it is difficult to establish the required links between thesauri and records disposal schedules, particularly when using automated records management systems.

Functional thesauri

There are a number of strengths and weaknesses of functionally structured thesauri, as summarised below:

Strengths

- seamless integration with common administrative terms thesauri – such as *Keyword AAA*
- seamless integration with *General Disposal Schedule 15* (State Records)
- seamless integration with future operational records disposal schedules, as State Records adopts a functional approach to schedule development.

Weaknesses

- can be complex to develop
- use language and structures that may not be generally familiar to agency staff
- require a considerable degree of initial research and preparation
- may not facilitate easy navigation to a particular subject.

Subject-based thesauri

There are also strengths and weaknesses in developing and using a class or subject-based thesaurus:

Strengths

- language and structures used are generally more familiar to agency staff who are not expert records managers
- easier to navigate for entries on specific subjects.

Weaknesses

- period of effectiveness/applicability can be limited
- prone to being very lengthy documents to include all of the subjects desired by the agency
- will not align with GDS 15 and future operational disposal schedules developed by State Records using a functional approach
- records of mixed retention periods may be placed on the same file creating difficulty in sentencing.

Recommendation

State Records recommends that, as a norm, South Australian Government agencies adopt a functional approach to the development of thesauri. Development of subject-based thesauri should be treated as the exception to the rule. If an agency chooses to develop a subject-based thesaurus it should contact State Records before proceeding with the project.

Part 3 – Merging of agency specific and general administrative thesauri

Explanation

Once the project team has established the existence of any thesauri of general administrative terms it should decide if the agency specific thesaurus should be merged with it. If the team decides to combine the two, the merged document needs to be carefully proof read to identify:

- spelling mistakes
- formatting mistakes
- dead end entries
- repetition of entries
- poor or inadequate cross-referencing.

Two thesauri can only be merged if both have been developed with identical structures. Any attempt to merge *Keyword AAA* with a class or subject-based thesaurus will fail.

Part 4 – Linkages between thesauri and disposal schedules

Explanation

There are a number of advantages for your agency in developing and using functional disposal schedules and thesauri:

- consistency in the use of language and terminology for both controlling and sentencing records
- easily transferable staff training
- the business classification scheme can be used as the basis for both tools
- the common functional keywords and activity descriptors help the appraisal and sentencing of a record at the point of creation or registration.

Where an agency uses records or document management software the import of functional disposal schedules and thesauri makes it possible to automatically assign a sentence to a record at the point of registration, saving time and reducing the possibility for sentencing errors.

In sentencing at creation or registration, however, your agency needs to ensure that any such automatically assigned sentences are reviewed by an experienced records management officer prior to the disposal of any records. This will ensure that where changes in content or function of a record occur after applying descriptive terms from the thesaurus, the previously assigned disposal action can be altered if necessary.

If a thesaurus is imported into records or document management software, compatibility issues concerning the software and the import format of the thesaurus needs to be addressed.

Glossary

State Records has produced an extensive Glossary of Records Management Terms. This can be accessed and downloaded from the Adequate Records Management, publications section of the State Records website, http://www.archives.sa.gov.au