

NSW Natural Resources Data Framework Thesaurus Development Final Report

Prepared for

**NSW Natural Resources Information Management
Strategy (NRIMS)**

Version	0.3
Status	Draft
Date	29 July, 2002
Job No.	17044

TABLE OF CONTENTS

1	Brief.....	1
2	Methodology.....	3
2.1	Begin a new thesaurus.....	3
2.2	Adapt an existing thesaurus.....	3
2.2.1	Candidates – GEMET.....	5
2.2.2	Candidates – CERES.....	5
2.2.3	Candidates – Environment Australia.....	7
3	The proposed thesaurus.....	8
3.1	Adapting GEMET.....	8
3.2	Thesaurus listing.....	10
3.3	Using the Thesaurus.....	11
3.4	Indexing a data set.....	11
3.5	Searching a data set.....	11
4	Candidate Source Reviews.....	14
4.1	Candidate terms lists.....	14
4.2	Classification of Natural Resource Terms.....	16
4.3	Review summaries.....	17
4.3.1	ANZECC Core Environmental Indicators (Classification).....	17
4.3.2	ANZLIC Metadata Guidelines (Terms List and Classification).....	18
4.3.3	CANRI Web site (Classification).....	18
4.3.4	CERES (Terms List).....	18
4.3.5	DLWC Glossaries (Terms List).....	19
4.3.6	Environment Australia Thesaurus (Terms List).....	19
4.3.7	EPA SoE Web site terms (Terms List and Classification).....	20
4.3.8	Global Spatial Data Infrastructure (Classification).....	20
4.3.9	Local Government SoE reporting guidelines (Classification).....	20
4.3.10	Manly Environmental Centre (Terms List and Classification).....	20
4.3.11	NAL Agricultural Thesaurus (USA) (Terms List).....	21
4.3.12	Natural Resource Management Ministerial Committee Framework _ national level Indicators (Classification).....	21
4.3.13	NLWRA Australian Natural Resource Atlas (Classification).....	21
4.3.14	NSW Catchment Blueprints template (Classification).....	21
4.3.15	EPA SoE 2003 Report (Classification).....	22
4.3.16	NSW EPA Thesaurus (Terms List).....	22
4.3.17	NT Office of Supervising Scientist Thesaurus (Terms List).....	22
4.3.18	Service NSW web sites (Classification).....	22
4.3.19	SOPA - GEMET thesaurus (Terms List).....	22
4.3.20	US Library of Congress (Terms List).....	23
4.3.21	WISE (DLWC) (Terms List).....	23
5	CANRI Data Set Classification.....	24
6	Maintenance of the Thesaurus.....	33
7	Realising Benefits from the Thesaurus.....	35
8	Thesaurus Software Review.....	38
8.1	Web references for thesaurus software.....	38
8.1.1	Australian Society of Indexers.....	38
8.1.2	Michael Middleton (Queensland University of Technology).....	38
8.1.3	American Society of Indexers.....	38

8.1.4	Willpower	39
8.1.5	State Recordkeeping, New South Wales	39
8.2	Thesaurus software available	39
8.3	Thesaurus software requirements	39
8.4	Term Tree 2000	41
8.4.1	Supplier	41
8.4.2	Notes	41
8.4.3	Import formats.....	41
8.4.4	Cost	41
8.5	MultiTes	42
8.5.1	Supplier	42
8.5.2	Notes	43
8.5.3	Import formats.....	43
8.5.4	Cost	43
8.6	WebChoir	44
8.6.1	Software available	44
8.6.2	Supplier	45
8.6.3	Notes	45
8.6.4	Cost	45
8.7	Oracle Text (formerly <i>interMedia</i> Text)	46
8.7.1	Software available	46
8.7.2	Supplier	46
8.7.3	Cost	46
9	Draft thesaurus Database	47
9.1	MS Access database	47
9.2	Terms database	47
9.2.1	Database tables	47
9.2.2	Attributes	49
9.2.3	Data models	50
10	References.....	52
11	Appendix 1 – Review of Terms Lists.....	53
12	Appendix 2 – Review of Classifications.....	54
13	Appendix 3 – CANRI Core and Priority Data Set Classification	55

DOCUMENT VERSION CONTROL

Title	NSW Natural Resources Data Framework Thesaurus Development
Document Status	Draft
Version	0.3
Date	29 July, 2002
Filename	C:_projects 2002\17044 NSW NRE Thesaurus\Reports\CANRI_Data_Framework_Thesaurus_Final_Report_v03.doc
Author(s)	D.R. Miller, R.G. Hume
Distribution	Data Framework Committee
Job No	17044

DOCUMENT HISTORY

Version	Comments	Date Issued
0.1	Initial draft	17 July 2002
0.2	Add classification mapping	28 July 2002
0.3	Revise chapter order	29 July 2002

DOCUMENT APPROVAL

Title	NSW Natural Resources Data Framework Thesaurus Development
Version	0.3
Approval	
Position	
Date	

1 BRIEF

The Community Access to Natural Resources Information (CANRI) program provides information products tailored for community-based local and regional environmental management in New South Wales. It is the first program of its kind to offer integrated access to maps and other data held at various Internet sites, by a diverse group of natural resource agencies and other stakeholders. CANRI is a collaborative initiative involving all NSW natural resource agencies.

The Data Framework project is funded by CANRI during 2001-2002 and overseen by the NRIMS Steering Group. The project will provide for the first time an overall structure for natural resources management datasets, endorsed by the key NSW natural resource management agencies. This high-level Data Framework will be published and reviewed by stakeholders. It will be used to create:

- a classification system for data served via CANRI;
- a taxonomy of natural resources terms for use in the CANRI catalogue, used by applications such as the Natural Resources Data Directory (NRDD) and Natural Resources Atlas (NRA);
- an enhanced CANRI homepage and map data selection tool allowing discovery of information within the branches of the Data Framework; and
- an XML schema to which more detailed schemas for individual data themes can be fitted as these become available.

The purpose of the consultancy is to develop a thematic structure for natural resources and environmental (NR&E) information across the NSW government. This structure will be submitted for formal endorsement by the natural resources management agencies under the framework of the Natural Resources Information Management Strategy (NRIMS).

In particular, the brief refers to the following outputs:

1. A master thesaurus. The listing for the master thesaurus draft is presented in Appendix 1
2. Mapping of the key existing classifications systems to the master thesaurus. The mapping of some 12 classifications is given in Appendix 2.
3. Mapping of the master thesaurus to a general purpose end user vocabulary. This was completed as part of the above, in this case, for the vocabulary developed by the Manly Environment Centre. This is shown in Appendix 2.
4. Recommend processes for maintaining and extending the thesaurus. This is detailed in Section 10 of the report
5. A classification, using the master thesaurus, of core and priority datasets. Classification of the data sets listed under the CANRI web site, and for which ANZLIC subject keywords are published, has been completed and is reported in Appendix 3.
6. Worked examples – classify a piece of NR&E information. This is contained in the core and priority data set classification (Appendix 3) with a further worked example in the text.
7. Worked example – harmonise a new classification system. This is contained in the classification review in Appendix 2.
8. Browse to find information. This is discussed in Section 8.2

9. Recommendations for realising benefits. This is discussed in Section 14.
10. Discussion of other significant issues. This is contained within the text.

2 METHODOLOGY

2.1 Begin a new thesaurus

Developing a new thesaurus for the NSW Natural Resources and Environment agencies would require the establishment of a number of subject matter expert committees. These committees would each be charged with developing the list of terms and the relationships between those terms for their particular subject area.

This was deemed an inappropriate development methodology as:

- There was insufficient time, resources and funds available to undertake such a process,
- Subject matter areas would have to involve representatives from more than one agency and this would be difficult to manage, and
- The merging of the subject matter areas would itself require a committee.

2.2 Adapt an existing thesaurus

In the last few years a number of Australian and international agencies have published thesauri in their particular areas. An exhaustive list produced by another recent consultancy undertaken for the NSW Office of Information Technology (OIT) by the State Library of NSW in February 2000, has listed some 12 thesauri that could be considered as candidates for an natural resources and environment thesaurus (Reference 1). That list includes:

Category	Title	Published
ENVIRONMENT	Acomplis thesaurus [microform].	[London : Greater London Council Research Library], 1 985
	Australian geoscience, minerals and petroleum thesaurus 1 Australian Mineral Foundation.4th ed.	Glenside, S. Aust. : Australian Mineral Foundation, c 1996.
	Californian Environmental Resources Agency - CERES thematic thesaurus.	http://www.ceres.ca.aov/technicalinfo/thesaurus.html
	Dictionary & thesaurus of environment, health & safety / U.S. Department of Energy, Office of Environment, Safety and Health.	Boca Raton, Fla.: C.K. Smoley, 1992
	Environmental thesaurus / compiled by B.E. Easterbrook	Bell Bay [Tas.]: Comalco, 1981
	EnVoc multilingual thesaurus of environmental terms.4 th ed.	Infoterra, Nairobi United Nations Environment Programme, 1997. http://www.cedar.univie.ac.at/une/p/infoterralthesaurus/
	SAGE : science and geography, education thesaurus 1 Compiled by CSIRO Information Services	East Melbourne: CSIRO Australia, c1995.
Australasian Spatial Date: 29 July, 2002	Data Exchange Centre	Page 3

NSW Natural Resources Data Framework Thesaurus Development

Category	Title	Published
	Thesaurus of environmental protection terms: subject terms for the West Australian Department of Environmental Protection developed by Kathy Little and Maggie Exon.	Little, Kathy. Perth : The Dept., c1995
RURAL	AGRI/AGRISANASTO: thesaurus of the Agricultural Library at the University of Helsinki.	http://wwwdb.heisinki.fi/triphomelacirisanasto/Welcomeng.html
	Agrovoc : multilingual agricultural thesaurus : English version. 2nd ed.	Rome: Apimondia by arrangement with the Food and Agricultural Organization of the United Nations, 1992. http://www.fao.org/scripts/aarovoc/frame.htm
	Agterms thesaurus for Australian agricultural extension Information.	Rural Industries Research and Development Corporation (Australia). Cloverdale, W.A. : Kondinin Group, 1999.
	CAB thesaurus. 1995 ed	Wallingford, UK: CAB International, 1995.
TRANSPORT	ITS thesaurus: a controlled hierarchical vocabulary for indexing materials relating to intelligent transport systems /prepared by Seyem D. Petrites and Michael C. Kleiber.	http://www.nas.edu/trb/about/itsthes.html
	ROAD thesaurus / compiled by Vija Walz and Andrew Pentecost.	Vermont South, Vic. : Australian Road Research Board, 1995.
	Thesaurus of transport terms / [prepared jointly by the Systems and Information Branch and the Library of the Bureau of Transport Economics].	Canberra : The Branch, 1981
UTILITIES	DWR merged keyword thesaurus / compiled by Margaret Wyatt & Liza Kendall. 2nd. rev. ed.	New South Wales. Dept. of Water Resources. [Parramatta, N.S.W.1 : Water Resources, 1992.
	EPRI thesaurus.	Palo Alto, Calif. : Electric Power Research Institute, 1988.
	Gas abstracts thesaurus	.Chicago, Ill. : Institute of Gas Technology, 1992

In recent times an additional thesaurus, the Generalised Multilingual Environmental Thesaurus (GEMET), has been developed for the Catalog of Data Sources (CDS) of the European Environmental Agency.

The candidate list of thesauri selected for consideration are:

- GEMET
- CERES
- Environment Australia

2.2.1 Candidates – GEMET

The Generalised Multilingual Environmental Thesaurus (GEMET) was conceived as a general thesaurus aimed to define a common general language. GEMET was initially created by merging the thesauri from 6 major European thesauri, including Infoterra as described above. This has been supplemented by an extensive review by the US EPA to yield the current version. The resulting 5334 terms have been arranged in a classification scheme made up of 39 groups with each term arranged in a hierarchy of terms. GEMET follows the ISO norms for monolingual and multilingual thesauri. (Reference 2).

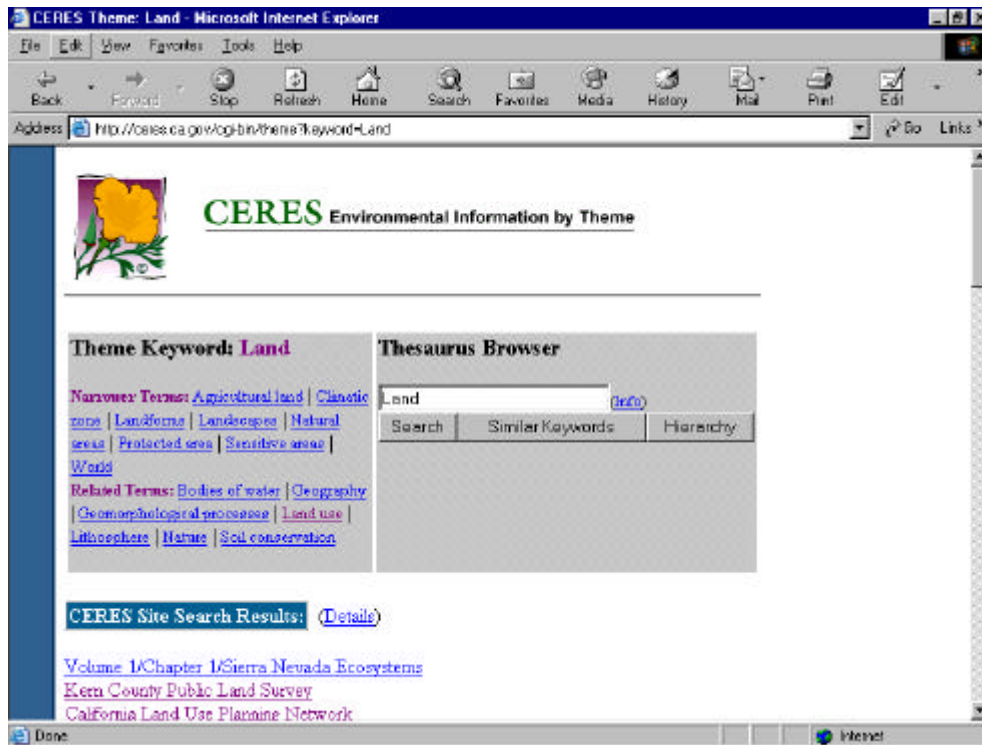
GEMET thesaurus is available as an Access database which can be freely modified. In addition, the EEA also distribute a program to show the thesaurus terms in alphabetic, thematic and systematic (hierarchical) views. This software, THESshow, is also freely available.

2.2.2 Candidates – CERES

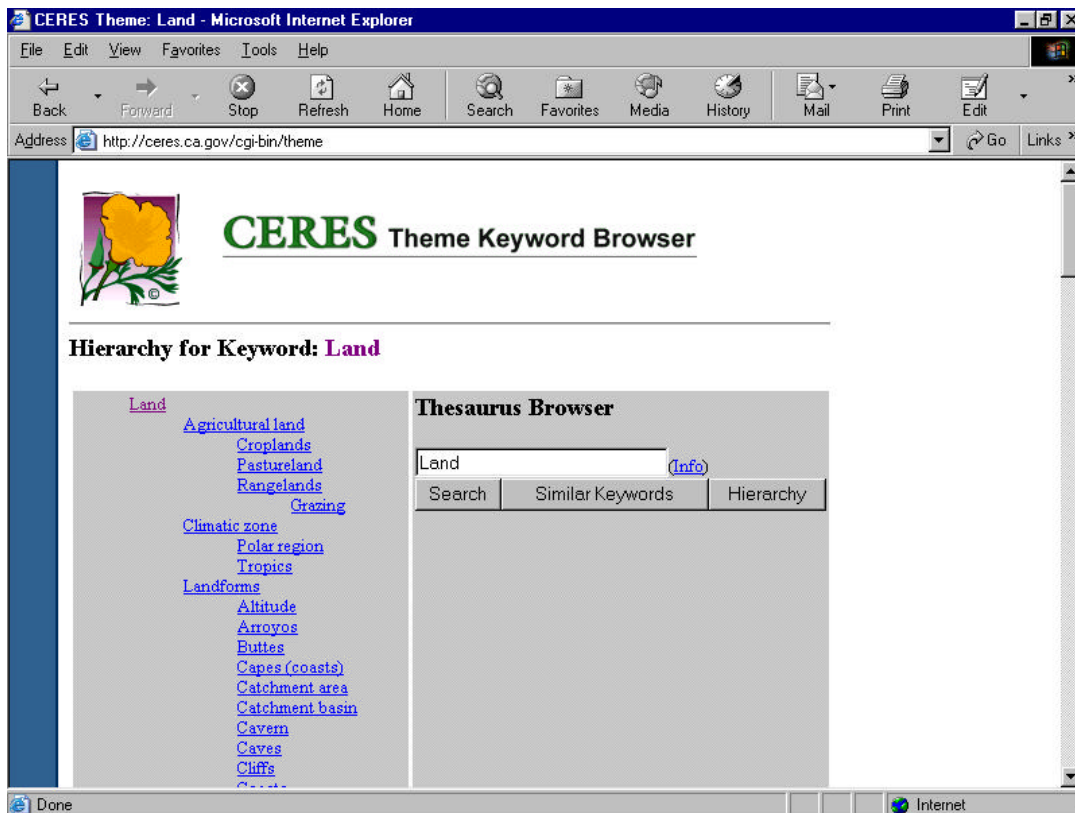
CERES is an information system developed by the California Research Agency to facilitate access to a variety of electronic data describing California's rich and diverse environments. The goal of CERES is to improve environmental analysis and planning by integrating natural and cultural resource information from multiple contributors and by making it available and useful to a wide variety of users. This thesaurus recognises two levels of classification, the top level with three terms:

- Natural Resources
- Natural Environment
- Human Environment

Below each of these levels there are up to 18 further classes of information. In total there are 1959 terms defined within the thesaurus. CERES presents the thesaurus with lists of broader terms, narrower terms and related terms for each term selected as a search terms, as shown below:



The terms can also be viewed in a hierarchy.



2.2.3 Candidates – Environment Australia

The EA Thesaurus is broadly based on a subset of the *Thesaurus of Environmental Protection Terms*, Department of Environmental Protection, WA, 1995. Over the years, current specific Commonwealth terminology has been incorporated into the thesaurus.

The thesaurus is small and comprises some 296 terms in total along with their broader term, narrower term and related term relationships. It is understood that this thesaurus is no longer in general use by Environment Australia.

3 THE PROPOSED THESAURUS

3.1 Adapting GEMET

On the basis of the comprehensive nature of GEMET, its acceptance in Europe and the USA, and the fact that it can be freely obtained and modified, the GEMET thesaurus was adopted as the starting point for the NSW Natural Resources Data Framework Thesaurus.

The structure of GEMET is such that there are 2 possible methods to impose the classification over the GEMET terms and their relationships. The first method makes use of the GEMET concept of groups. Within GEMET each top term is allocated to a group. That is the group points directly to a “top term” below which a set of terms is linked by the broader term/narrower term relationships. By linking a high level classification term to the top term all the words associated with that top term are automatically carried over to sit under the high level classification term.

The second method available is to use the GEMET concept of theme. GEMET allows each term to be linked to multiple themes. The concept allows a “thematic” view of all the terms in the thesaurus to be displayed as each term is linked to one or more themes. The limitation is that the terms under the theme can only be displayed alphabetically as the broader term/narrower term relationships are not used.

For the first version of the NSW Natural Resources Thesaurus an alternative set of “Groups” has been mapped to the GEMET top terms. This retains the hierarchical structure within GEMET.

The group terms available in GEMET have been rejected primarily because the terms are not in common usage (“Anthroposphere”). The alternative top level group classification used conforms to the mandatory requirements of the International Standard ISO 19115 – Geographic Information - Metadata . This standard defines a mandatory term TopicCategory and mandates the use of 19 terms to describe geographic data. These topic categories are given in the following table.

topicCategoryI	topicCategory
001	Farming
002	Biota
003	Boundaries
004	Climatology/Meteorolo
005	Economy
006	Elevation
007	Environment
008	Geoscientific
009	Health
010	Imagery/Base
011	Intelligence/Military
012	Inland Waters
013	Location
014	Oceans
015	Planning/Cadaastre
016	Society
017	Structure
018	Transportation
019	Utilities/Communication

In Australia there is currently a draft profile of this standard which will eventually become the 3rd version of the ANZLIC metadata guidelines. For this reason the mandatory elements of the TopicCategory field have been taken from the standard and superimposed over GEMET by mapping those terms to the GEMET Top Term. By adopting this methodology whole branches of terms, as related within GEMET by Broader and Narrower Term relationships, are made available to the proposed thesaurus. This methodology thus utilises the resources that have gone into the GEMET terms relationships along with conformance to a developing International Standard.

The ANZLIC Metadata Guidelines contain core metadata elements for cataloguing geographic datasets in Australia and New Zealand. The guidelines also recommend keywords for cataloguing, called 'Search Words'. Associated with the 153 Search Words are 23 Qualifier words, which can be added to the Search Words. ANZLIC representatives have been closely associated with the current development of the ISO Standard for geographic information metadata, ISO/DIS 19115 Geographic information – Metadata and intend to harmonise the guidelines with this standard.

AUSDEC have discussed the harmonisation with Craig Macauley who has represented Standards Australia and ANZLIC on the ISO TC211 metadata committee to obtain an indication of the direction that ANZLIC will take in adopting the ISO standard.

The standard is currently a Draft International Standard, which implies that it is in near final form and is expected to be ratified towards the end of this year. Craig has completed a draft ANZLIC profile which has been recently published for comment at the ANZLIC web site which proposes a mapping between the ANZLIC schema and the ISO standard for the metadata elements, that is for elements such as 'Title', 'Custodian', 'Abstract'.

The ISO Standard is written using the Unified Modelling Language (UML) to rigorously describe the metadata. It defines mandatory and conditional metadata sections, metadata entities and metadata elements. The standard defines core metadata for geographic metadata.

Within the core metadata are 19 dataset topic categories to be used to describe the datasets at a high level. The topic categories include such terms as farming, biota, boundaries and Climatology/Meteorology/Atmosphere. The ISO Standard also includes provision for keywords to enable further categorisation of datasets. These keywords can be adopted from a named thesaurus and function at a lower level to the topic categories. This enables keywords to be used from both the ANZLIC Metadata Guidelines and from GEMET.

An ANZLIC project team is to be formed to review the profile and later is expected to include a review of the keywords. There is no indication yet of what will be done with the ANZLIC search words but it has been suggested that the ANZLIC search words could be replaced with ISO topic categories.

AUSDEC do not believe that the 19 ISO topic categories can replace the 153 ANZLIC search words because the search words offer greater discrimination. The topic categories group datasets into broad categories such as environment and economy. The ANZLIC search words contain 30 broad categories such as 'AGRICULTURE' but provide further discrimination within these categories such as 'AGRICULTURE Crops' and 'AGRICULTURE Irrigation'. While many of the ANZLIC broader categories map directly to the ISO topic categories such as 'AGRICULTURE' to 'farming' others do not such as 'FORESTS' is contained within 'farming' and 'economy'.

For this current project, AUSDEC have mapped the majority of the 153 ANZLIC search words to GEMET thesaurus terms. (No consideration was given to mapping the ANZLIC Qualifier terms because these are used to qualify the search words and the combination of the terms becomes too unwieldy.) The GEMET thesaurus terms are networked in a hierarchy (polyhierarchical thesaurus). AUSDEC have mapped the 19 ISO topic categories to the 39 top terms in the GEMET hierarchy so that there is a relationship between the ANZLIC topic categories and the ANZLIC search words through the hierarchy of GEMET terms.

3.2 Thesaurus listing

The proposed thesaurus is presented as a Word file and a text listing (Appendix 1). In the text listing each level of the thesaurus hierarchy is represented by a (.). Thus the full hierarchy for the term "national park" is presented as:

```
Environment  
.land  
..protected area  
...national park
```

The Word version has heading levels set for each of the levels and may be viewed using the View/Outline menu. In this view each heading level can be switched on and off and the thesaurus can be negotiated by selecting more and more levels. Thus selecting heading 1 shows the TopicCategories, selecting level 2 shows the TopicCategories and the Top Terms

and so on. There are up to 10 levels of hierarchy in the thesaurus. Selecting all levels to show will yield the text listing as appended.

3.3 Using the Thesaurus

The thesaurus serves two major purposes:

- As a glossary of terms that may be used to index data sets,
- As a navigation tool to assist users to see data sets classified using the vocabulary.

These uses are illustrated below.

3.4 Indexing a data set

A natural resource data set can be indexed using the vocabulary provided by the thesaurus. The current draft International Standard ISO 19115 – Geographic Information- Metadata requires that each data set be indexed by at least the TopicCategory terms defined in the thesaurus. This requirement is carried over into the draft ANZL:IC metadata guidelines currently under review.

For example, a data set titled “ Soil Maps on NSW” could be indexed as follows:

1. Currently the data set could be indexed using the ANZLIC Keywords
 GEOSCIENCES Geology
 LAND Cover
 SOIL
2. The thesaurus can provide the following terms
 Geology(Geoscientific Information.science.natural science.earth science)
 Land Cover (Environment.environmental data)
 Soil map (Society.information.documentation.document.document type)
 Soil Profile (Geoscientific Information.pedosphere.soil)

The data set would be registered with each word in the hierarchy (as shown in the brackets). This could be an automatic process with custom software that selects all terms higher in the hierarchy than the selected term and adds it to the list of index words for the data set.

3.5 Searching a data set

There are three methods to search for a data set. All revolve around using the index words associated with the data set.

In the first method the user is asked for a search term and this is used. This method provides the user with no assistance in selecting a term and assumes they are familiar with the terminology. This is not a recommended method for searching.

The second method provides the user with access to the thesaurus hierarchy. In this case the user is presented with the TopicCategories as the first level of searching:

- Biota
- Climatology/Meteorology/Atmosphere
- Economy
- Environment
- Farming
- Geoscientific Information
- Health
- Imagery/Base Maps/Earth Cover
- Inland Waters
- Planning/Cadaastre
- Society
- Structure
- Transportation
- Utilities/Communication

Once one is selected, “Geoscientific Information” for example, then all the top terms available under that term are displayed, in this case:

- Analysis
- Environmental data
- Evaluation
- Experiment
- Geological process
- Lithosphere
- Methodology
- Monitoring
- Parameter
- Pedosphere
- Research
- Science
- Statistics

Selecting “pedosphere” then displays two further terms:

- Soil
- Soil process

Selecting “soil” then displays eight more terms:

- Soil air
- Soil function
- Soil profile
- Soil salinity
- Soil structure
- Soil texture
- Soil type
- Soil water

Selecting “soil profile” displays the data set in question.

NSW Natural Resources Data Framework Thesaurus Development

At each stage in the process all the data sets that have the index words selected could be displayed. Thus under the first selection “Geoscientific Information” there could be thousands, under the next term “Pedosphere” hundreds, and so on.

The final method of searching involves a pre-defined template of terms from the thesaurus. The terms used in the template do not need to be in their correct hierarchy and do not need to include all terms. For instance, the top level of terms presented could list say only 10 of the 19 TopicCategories. In this way data sets on “Health”, “Society”, “Utilities” etc could be removed from the view of the data sets that is being presented. Similarly if a level below is shown in the template it too could select only those terms that the web administrator wishes to provide access to.

For instance a predominantly environmental web site could present a structure as follows:

Biota		
Biochemical process		
Environment		
Environment	assessment	control
impact	management	Planning
Policy	Problem solving	Protection
Pollution and Waste		
Pollution	Pollutants	Waste

Each of these terms would display all data sets that included these index terms and those below them in the hierarchy.

4 CANDIDATE SOURCE REVIEWS

The review of the NSW Natural Resource Thesaurus was undertaken in two stages:

- Review the terms in use in the natural resource and environment community
- Review the classification of the terms in groups and themes as a method of presenting the terms in a logical hierarchy

4.1 Candidate terms lists

The first task undertaken was to compare the terms used in NSW natural resource and environment information sources and selected Commonwealth and International sources with those already contained within the GEMET thesaurus. The lists of terms chosen show the terms in use in the community and where a match occurs that term satisfies the criteria of common usage and existence within GEMET, meaning that the term could be used as an indexing term.

The review considered 12 sources of terms, including both lists of terms reflecting NSW and Commonwealth natural resource terminology and international terminology. The source lists are summarised in the table below. This shows, for each source list:

- the number of terms,
- the number of direct matches to GEMET,
- the number of matches where an alternate term was chosen, and
- the number of terms where no match was found.

Alternate terms were chosen to remove plurals (sewage systems to sewage system), change of word order (state of environment report to report on the state of the environment) and local language (aeroplane to aircraft). However others reflect a synonym choice (marine habitat to marine ecosystem).

Where a direct correspondence exists the term could be used as an index term for a data set. Where an alternate term has been proposed it is recommended that subject matter experts validate the alternate term. Where no matching term is proposed then that term could be promoted for inclusion in the thesaurus. The addition of such terms delineates the difference between the GEMET and the NSW Natural Resources Thesaurus.

NSW Natural Resources Data Framework Thesaurus Development

The candidate terms lists reviewed included:

Source of Candidate Terms	Description	No Terms	GEMET match	Alternate term match	No match
ANZLIC Metadata Guidelines	Subject keywords	153	27	116	9
CERES	Californian environmental thesaurus aligned with GEMET	1959	900	64	995
DLWC Glossaries	Internal information management system	520	89	115	314
Environment Australia Thesaurus		296	100	148	48
EPA SoE Web site terms	taken from web site classification	72	17	40	15
Manly Environmental Centre	Internal classification	203	38	71	94
NAL Agricultural Thesaurus (USA)	US Agricultural thesaurus	746	138	8	600
NSW EPA Thesaurus		1119	159	251	709
NT Office of Supervising Scientist Thesaurus		185	39	28	118
SOPA Gemet thesaurus as revised	Sydney Olympic Park Authority thesaurus developed from GEMET and extended for remediation terms	6394	5781	0	613
US Library of Congress	Partial listing for ecology and nature conservation	118	16	1	101
WISE (DLWC)	DLWC Water Information System CD	301	65	169	67

A short summary of each of the candidate term lists is given later in this section. The candidate terms lists and their correspondence in GEMET are given in Appendix 1 in a series of Excel spreadsheets, one for each source list. The spreadsheet gives the candidate term, any alternate term generated, the GEMET term, and the hierarchy for that term up to the Top Term in GEMET. The term hierarchy is important in the further consideration of the term as part of a classification of natural resource terms. This is discussed in further detail below.

The review shows that of the over 10,000 terms reviewed, that the GEMET thesaurus had matches for all but 3683 (36%). For those terms sourced from NSW and the Commonwealth (2849 terms) all but 1374 (48%) found matches in GEMET. The review suggests that of these about 20% are considered natural resource terms. Thus there is a potential to promote around 300 to 500 terms to the thesaurus to accommodate NSW and Commonwealth usage. On this basis it is recommended that the GEMET thesaurus be adopted as the first version of the NW Natural Resources Thesaurus.

It is recommended that the terms lists developed now be reviewed, by subject matter experts, to determine whether new terms need to be added to accommodate the terms where no match was found and whether the choice of alternate term made by the project team is acceptable.

4.2 Classification of Natural Resource Terms

An important aspect of the development of the NSW Natural Resource Thesaurus is the development of a classification of the terms contained within the thesaurus.

The purpose of a classification is to provide a way to present information sources that are indexed under the terms (and the hierarchy) embodied in the thesaurus. There are three ways to do this:

1. Present a limited or selective view of the hierarchy. This is the view currently represented by the CANRI web site. It shows 7 themes and for each one a number of data sets that are classified under that theme.
2. Present the complete hierarchy allowing the user to move up and down the hierarchy using the broader term/narrower term relationships. This is represented by the current CERES web site, which is illustrated earlier in this report.
3. Do not use the hierarchy at all and simply present an alphabetic list of the terms and let the user select a term(s) and see the information sources indexed to it. This is represented by the “search term” query box available on most web search engines. This class of view could also show information sources by agency or data type etc.

The classification provides a way to present natural resource data in a logical and familiar way. The classification must reflect both the coverage of the terms and the commonly used major thematic areas. The latter are a product of the time, for instance, some years ago biodiversity was considered an aspect of flora and fauna, now it enjoys prominence and is reflected in the higher levels of many classifications despite its overlap with other major thematic areas.

The various classification schemes have been reviewed by determining where the terms under each class map within the proposed classification. A well-formed mapping will have the following characteristics:

- Each class has a class in the proposed classification even if a different name.
- A similar level of classification – thus if Agriculture is the first level in a classification it should be a first, or at worst second level in the proposed classification, and
- All the terms classified under one class in the classification should map to one class in the proposed classification.

The review considered 12 classification sources, including both classifications reflecting NSW and Commonwealth natural resource terminology and international terminology. The classification source lists are summarised in the table below.

Source of Candidate Terms	Description
ANZECC Core Environmental Indicators	Taken from chapter headings from the SoE 2000 report
ANZLIC Metadata Guidelines	Subject keywords
CANRI web site	existing web site classification
EPA SoE Web site terms	taken from web site classification
GSDI	
Local Government SoE reporting Guidelines	
Manly Environmental Centre	Internal classification
Natural Resource Management Ministerial Committee Framework _ national level Indicators	
NLWRA Australian Natural Resource Atlas	taken from web site classification
NSW Catchment Management Blueprints	Proposed classification (nature.net)
NSW EPA SoE Chapter headings 1993-2003	Taken from chapter headings from all SoE reports since 1993 to 2003(in writing)
Service NSW web site	

A brief review of the mapping of the various classification schemes is given below. Details of each candidate classification and its mapping to the proposed thesaurus classification are given in Appendix 2.

Most classifications have mapped well to the proposed classification in that the particular theme in the classification list has a corresponding term in the thesaurus that is classified under a top term. Many classes in the lists, however, map to multiple Groups in the proposed classification. Thus the grouping used in the candidate classification is lost under the proposed classification. This is a minor problem as the mapping can show the relationship of one to the other and a query developed under a candidate system could be mapped to the appropriate classes under the proposed classification. For instance, if a user were to select “Atmosphere.stratographic ozone depletion.concentration of ozone depleting substances” from the EPA 2003 classification this would map directly to “Atmosphere.effect” in the proposed classification. In this case a query under this class would also yield this climatic effect as well as biological, environmental and landscape effects. This is a much broader query.

4.3 Review summaries

4.3.1 ANZECC Core Environmental Indicators (Classification)

The Australia New Zealand Environment and Conservation Council (ANZECC) has developed a set of core environmental indicators for the reporting of the State of the

Environment. These are published in the March 2000 report of the Australia New Zealand Environment and Conservation Council State of the Environment Reporting Task Force. (Reference 3).. These indicators are presented in a classification of theme/Issue and core indicator.

This classification relates to much of the natural resource information to be accessed by CANRI and the classification is relevant to the proposed CANRI classification.

Overall the ANZECC classification maps easily to the proposed classification. The major issue is with the high level term “biodiversity”. This covers lower classifications relating to aquatic habitat, native vegetation, fire, introduced species and extinct, endangered and threatened species. All these terms lie under two TopicCategories – Biota and Environment. Similarly Human Settlements maps to a variety of TopicCategories.

4.3.2 ANZLIC Metadata Guidelines (Terms List and Classification)

ANZLIC have developed the ANZLIC metadata guidelines (Reference 4). Currently all natural resource information to be accessed by CANRI must have ANZLIC metadata to these guidelines. Within the guidelines are a set of subject keywords. These keywords represent a classification as well as a controlled vocabulary and provide a list of terms. This list comprises some 153 terms grouped under 30 themes. Both the terms and the themes have been used to generate the candidate terms, for instance, “ECOLOGY Community” has been matched to “Ecological Community”. The terms are both single word and phrases. This explains the high number of alternate terms used.

The majority of search terms match GEMET terms with only 9 terms having no direct match. This includes terms relating to “native” or “aboriginal” and some very specific terms like “El Nino”. The general terms for native, indigenous and aboriginal all need to be promoted to the thesaurus.

The ANZLIC subject keywords include a simple single level classification comprising some 31 classes. The classification is designated by the capitalised terms, for example, AGRICULTURE Crops. The first level terms have been map well to the proposed classification. Some terms point to lower levels in the thesaurus and the top term they are related to may appear remote, for example ECOLOGY maps to Biota . biosphere.

4.3.3 CANRI Web site (Classification)

The CANRI web site includes a simple single level classification comprising 7 classes (Reference 5). The CANRI classification maps well except for the class Natural Resources and Environment class. This maps to many classes in the proposed classification. This would be expected for such a large thematic domain.

4.3.4 CERES (Terms List)

The California Environmental Resources Evaluation System (CERES) is a comprehensive web site that includes a search engine that references a environmental thesaurus (reference 7). This thesaurus has used the GEMET thesaurus in its development. The Terms list from CERES comprises 1959 terms. It is organised under a two level classification. CERES references GEMET in its background information and 964 terms match between the two

thesauri. Of the remaining terms most mismatches occur as CERES uses the plural form whilst GEMET uses the singular. Thus “Algal blooms” and “Algal bloom” do not currently match. In addition US spelling is used (“industrialisation” instead of “industrialisation”) A review of the terms list to define alternate terms would be expected to find simple matches for around 80% of the currently mismatched terms. This was not undertaken as the thesaurus is not Australian and as such was not a high priority.

The CERES thesaurus also includes a classification of terms in two levels. This classification has not been considered in this review.

4.3.5 DLWC Glossaries (Terms List)

The DLWC thesaurus will be used to classify all Departmental records, paper or electronic (Reference 8). This thesaurus is being developed under the KeywordsAAA guidelines developed by NSW State Archives. This provides a “function and activity” thesaurus not a “subject” thesaurus. This is a fundamental difference from that proposed for the CANRI thesaurus which by nature is a subject thesaurus.

The DLWC has made available a glossary of terms being used to classify DLWC information sources. This comprises some 520 terms and the terms are either subject or activity descriptors. The majority (314) of these terms do not match with GEMET terms. This is an important glossary. As found for the ANZECC list all references to “native”, “aboriginal” or “indigenous” are missing from GEMET and should be promoted. The remaining terms include a large number of government and business terms like, “irrigation district” “local environmental plan”. These terms will need to be considered in the thesaurus. Finally, the review indicates that GEMET is not comprehensive for terms relating to surface water management. This is a major DLWC activity and terms like “channel”, “river gauge”, “pumping”, “bores” etc should be promoted to the thesaurus.

4.3.6 Environment Australia Thesaurus (Terms List)

The EA Thesaurus (Reference 9) is broadly based on a subset of the *Thesaurus of Environmental Protection Terms*, Department of Environmental Protection, WA, 1995. Over the years, current specific Commonwealth terminology has been incorporated into the thesaurus. It is understood that this thesaurus is no longer in general use by Environment Australia, nonetheless it is a useful list of terms in current use.

The thesaurus is small and comprises some 296 terms in total. GEMET contains all but 48 terms. Many of these names are place names (“Queensland”), it is recommended that these terms should not be promoted to the thesaurus as gazetteers are designed to provide glossary lists of valid locations. There are however many terms which genuinely need promoting that related to Australian usage, like, “old growth forest”, “remnant vegetation”, etc.

Recently EA indicated that this thesaurus is no longer in use and that a whole of government thesaurus is being developed to replace it. This is not an issue for this review as the existing thesaurus is still a valid glossary of natural resource terms.

4.3.7 EPA SoE Web site terms (Terms List and Classification)

The EPA web site lists some 72 indicators under 6 classes (Reference 10). The list was taken from the State of Environment Report 2000 and corresponds to the major SoE indicators used for reporting at that time. All but 15 of the terms have matching GEMET terms. This reflects the strong environmental background of GEMET.

There are 6 classes in the first level of the SoE web site classification. This is then broken down to a further 72 indicators. Mapping of the classes shows that for some classes the mapping, whilst complete, distributes a single class in the original to many classes in the proposed thesaurus classification. In particular the components of the class “environmental sustainability” maps to at least 9 classes in the proposed classification. This is to be expected for such a high level class and is also seen in the “Human settlement” class. Some classes do not have a matching class in the current version of the thesaurus., for example “green products”. This would require the addition of new terms to the thesaurus, which could then be mapped into the classification.

4.3.8 Global Spatial Data Infrastructure (Classification)

The Global Spatial Data Infrastructure (GSDI) supports ready global access to geographic information (Reference 11). This is achieved through the coordinated actions of nations and organisations that promote awareness and implementation of complimentary policies, common standards and effective mechanisms for the development and availability of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes. These actions encompass the policies, organisational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their objectives.

The GSDI web site presents spatial data in a classification of some 20 classes. These map well to the proposed classification.

4.3.9 Local Government SoE reporting guidelines (Classification)

The State of the Environment Reporting undertaken by Local Government in NSW is governed by published environmental guidelines (Reference 12). These guidelines comprises 9 major classification classes and 54 topics. The classes each map to the proposed classification but generally are split up into a number of classes in the proposed classification . This applies particularly to the “Air”, “Land” and “Biodiversity” classes. One class relating to licensing does not have an corresponding class in the proposed classification.

4.3.10 Manly Environmental Centre (Terms List and Classification)

The Manly Environmental Centre provides a catalog of environmental reference materials (Reference 13). It has been developed over some time and has a focus on providing a classification of the information that is publicly accessible and comprehensible. It has a high level classification of some 12 terms under which some 203 terms are grouped. These terms are pitched at a community level. Some 94 terms have no match in GEMET. Many of these reflect locality names (“Rousse Hill”), Agency names, library references or information types (magazines etc).

The Manly Environmental Center also provides a classification system. This has been developed for public use and reflects public terminology, for example “Green Business”. The classification does map to the proposed classification but in this case the major classes do not correspond to the major classes in the proposed classification. Thus whilst most of the classes map across they are broken up. For example, the GREENHOUSE class is broken up into 7 classes in the proposed classification. Thus this mapping is not well reflected in the proposed classification.

4.3.11 NAL Agricultural Thesaurus (USA) (Terms List)

The NAL has developed and published an Agricultural Thesaurus (Reference 14). This was developed for anyone describing, organising, and classifying agricultural resources.

The thesaurus includes a classification of some 17 “subject categories”, one of which is “Natural resources, Earth and Environmental Sciences”. This subject category alone has some 746 terms. These terms mapped very poorly to GEMET with some 600 finding no match. As this is an overseas source manually defining alternate terms for plurals etc was not undertaken.

4.3.12 Natural Resource Management Ministerial Committee Framework _ national level Indicators (Classification)

The NRM has published a framework for reporting national level environmental indicators (standards & targets and monitoring & evaluation) (Reference 15). A review of these yields some 4 themes and 17 indicators. The terms do contain multiple terms.

The classes in the NRM framework map directly to classes in the proposed thesaurus classification. However a number of classes will map to two classes in the proposed classification. This is particularly true for the classes under the “Biodiversity” theme. The “Salinity and Water Quality Flows” theme maps to many classes. In this case the theme relates to the effect of the indicator as it relates to the theme, so the classification separates out for “native vegetation” effects under both Salinity and Biodiversity themes. The proposed classification can not accommodate this easily.

4.3.13 NLWRA Australian Natural Resource Atlas (Classification)

The NLWRA has developed a web site – the Australian Natural Resources Atlas (reference 16). The web site gives a classification of natural resource information with 11 themes and some 156 terms.

The classification maps well to the proposed classification with the exception of the “rangelands” classification. There is no corresponding class in the proposed classification for this. It could be accommodated by defining a new term under the “land” top term.

4.3.14 NSW Catchment Blueprints template (Classification)

There are many common strategic targets which are currently listed inconsistently across Catchment Blueprints in NSW (Reference 17)). These Catchment Blueprints are submitted for accreditation and funding and should be reporting to a consistent set of targets. The group

has suggested a classification comprising some 17 classes. This classification does not map well into the new classification. Whilst all but 3 classes are accommodated (and these classes are strictly not natural resource terms like “critical assets”) the remainder map to two classes “Geoscientific Information.science” and “Biota.biosphere”.

4.3.15 EPA SoE 2003 Report (Classification)

The NSW EPA has produced, since 1993, a series of reports on the State of the Environment (Reference 18). Each report has classified environmental data into a series of themes, issues and indicators. These have been successively refined over each report. The chapter headings for the 2003 NSW State of the Environment Report have been provided by the EPA. These yield some 6 themes with a further 71 sub classes.

The classification maps into the proposed classification quite well. The major themes map to one or two of the TopicCategories in the proposed classification and each sub-theme is catered for. The exception to this are the “Human settlement” and “Biodiversity” themes which maps to 7 or 8 classes in the proposed classification.

4.3.16 NSW EPA Thesaurus (Terms List)

The NSW has published a thesaurus of some 1119 terms (Reference 19). This thesaurus has not matched well with GEMET with some 709 terms finding no match. This is a poor matching for an important source of terms for NSW. It is explained by a large number of management/office terms like “after hours number”, “CEO review” etc. A scan of the unmatched terms suggests that only about 20% of the unmatched terms are natural resource terms. Thus the match is better than suggested by the raw figures.

4.3.17 NT Office of Supervising Scientist Thesaurus (Terms List)

The Northern Territory Office of the Supervising Scientist has created a comprehensive thesaurus (Reference 20). The thesaurus includes term relationships (broader/narrower terms) and is used as a controlled vocabulary for the classification of all reports and information produced by the Office. The thesaurus comprises some 185 terms, of which 118 failed to find a match in GEMET. Many of these terms are locality and company/agency names and are not relevant.

4.3.18 Service NSW web sites (Classification)

The Service NSW web sites lists 7 themes respectively for the classification of natural resource data. (Reference 6) The classifications both point directly to data sets. The classification maps well to the proposed classification with the exception of the class “Flora and Fauna”. This maps to 3 classes in the proposed classification as would be expected for such a broad class.

4.3.19 SOPA - GEMET thesaurus (Terms List)

The Sydney Olympic Park Authority (SOPA) has developed a Thesaurus for the classification of the large amounts of information they have collected, particularly relating to the remediation of the Olympic Park site (Reference 21). They have adopted GEMET in total and added some 613 additional words that were not in the GEMET terms. These relate to

localities and company/agency names (some 200 terms) and remediation terms (400 terms) that SOPA found were not well represented in the original GEMET terms.

4.3.20 US Library of Congress (Terms List)

A short extract from the US Library of Congress web site has yielded some 118 terms (Reference 22). Of these 101 did not match GEMET terms. These terms relate to terms under the ecology theme suggesting this area is not well represented in GEMET. These terms should be considered for promotion to the thesaurus.

4.3.21 WISE (DLWC) (Terms List)

The WISE system includes a glossary of terms (Reference 24). This terms list comprises some 301 terms. This list was well matched by GEMET with only 67 terms failing to find a match. Included in these terms are acronyms and some specific terms like "flood plain connectivity". Subject matter experts will need to review these terms for possible promotion to the thesaurus.

5 CANRI DATA SET CLASSIFICATION

The 500 data sets listed under the CANRI web site as core and priority data sets have been classified under the proposed classification. This has been done by using the subject keywords provided for each data sets as part of the ANZLIC metadata listing. These are broad index terms and are of limited number (153). As such the classification represents only a first cut as the thesaurus now provides many more specific index terms that can be assigned to the data sets. The task of assigning these index terms lies with the data set owners.

The data set classification is presented in Appendix 3.

The basis of the classification is the mapping of ANZLIC subject keywords to the proposed TopicCategory in the Thesaurus. This has been established by matching the ANZLIC term to a GEMET term and then determining the TopicCategory that contains that term. This mapping is presented below for reference.

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
AGRICULTURE	agriculture	Environment	agriculture
AGRICULTURE	agriculture	Farming	agriculture
AGRICULTURE Crops	crop production	Environment	agriculture
AGRICULTURE Crops	crop production	Farming	agriculture
AGRICULTURE Horticulture	horticulture	Environment	agriculture
AGRICULTURE Horticulture	horticulture	Farming	agriculture
AGRICULTURE Irrigation	irrigation	Environment	agriculture
AGRICULTURE Irrigation	irrigation	Farming	agriculture
AGRICULTURE Livestock	livestock	Biota	biosphere
ATMOSPHERE	atmosphere	Climatology/Meteorology/Atmosphere	atmosphere
ATMOSPHERE Air Quality	air quality	Environment	environmental assessment
ATMOSPHERE Greenhouse	greenhouse effect	Biota	effect
ATMOSPHERE Greenhouse	greenhouse effect	Climatology/Meteorology/Atmosphere	effect
ATMOSPHERE Greenhouse	greenhouse effect	Environment	effect
ATMOSPHERE Ozone	ozone		chemical element
ATMOSPHERE Pressure	pressure		physical property
BOUNDARIES	border	Imagery/Base Maps/Earth Cover	border
BOUNDARIES Administrative	administrative boundary	Imagery/Base Maps/Earth Cover	border
BOUNDARIES Biophysical			
BOUNDARIES Cultural			
CLIMATE AND WEATHER	climate	Climatology/Meteorology/Atmosphere	climate
CLIMATE AND WEATHER Climate change	climatic change	Transportation	traffic
CLIMATE AND WEATHER Drought	drought	Climatology/Meteorology/Atmosphere	climate
CLIMATE AND WEATHER El Nino			

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
CLIMATE AND WEATHER Extreme weather events	weather	Climatology/Meteorology/Atmosphere	climate
CLIMATE AND WEATHER Meteorology	meteorology	Geoscientific Information	science
CLIMATE AND WEATHER Radiation	radiation		radiation
CLIMATE AND WEATHER Rainfall	rain	Climatology/Meteorology/Atmosphere	climate
CLIMATE AND WEATHER Temperature	temperature		physical property
DEMOGRAPHY	demography	Society	demography
DISEASE	disease	Health	disease
ECOLOGY	ecology	Biota	biosphere
ECOLOGY	ecology	Geoscientific Information	science
ECOLOGY Community	ecological community	Biota	biosphere
ECOLOGY Community	ecological community	Geoscientific Information	science
ECOLOGY Ecosystem	ecosystem	Biota	biosphere
ECOLOGY Ecosystem	ecosystem	Geoscientific Information	science
ECOLOGY Habitat	habitat	Biota	biosphere
ECOLOGY Habitat	habitat	Geoscientific Information	science
ECOLOGY Landscape	landscape ecology	Geoscientific Information	science
ENERGY	energy	Utilities/Communication	energy
ENERGY Coal	coal		product
ENERGY Electricity	electricity		physical property
ENERGY Petroleum	petroleum		material
ENERGY Renewable	renewable energy source	Utilities/Communication	energy
ENERGY Use	energy utilisation	Utilities/Communication	energy
FAUNA	fauna	Biota	biosphere

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
FAUNA Exotic	exotic species	Biota	biosphere
FAUNA Exotic	exotic species	Geoscientific Information	science
FAUNA Insects	insect	Biota	biosphere
FAUNA Invertebrates	invertebrate	Biota	biosphere
FAUNA Native			
FAUNA Vertebrates	vertebrate	Biota	biosphere
FISHERIES	fishery	Farming	fishery
FISHERIES Aquaculture	aquaculture	Farming	animal husbandry
FISHERIES Freshwater	inland fishery	Farming	fishery
FISHERIES Marine	marine fishery	Farming	fishery
FISHERIES Recreational			
FLORA	flora (biology)	Biota	biosphere
FLORA Exotic	exotic species	Biota	biosphere
FLORA Exotic	exotic species	Geoscientific Information	science
FLORA Native			
FORESTS	forest	Environment	land
FORESTS Agriforestry	agroforestry	Farming	forestry
FORESTS Natural	natural forest	Environment	land
FORESTS Plantation			
GEOSCIENCES	earth science	Geoscientific Information	science
GEOSCIENCES Geochemistry	earth science	Geoscientific Information	science
GEOSCIENCES Geology	geology	Geoscientific Information	science
GEOSCIENCES Geomorphology	geomorphology	Geoscientific Information	science
GEOSCIENCES Geophysics	geophysics	Geoscientific Information	science
GEOSCIENCES Hydrogeology	hydrogeology	Geoscientific Information	science
HAZARDS	hazard	Health	hazard
HAZARDS Cyclones	cyclone	Climatology/Meteorology/Atmosphere	climate
HAZARDS Drought	drought	Climatology/Meteorology/Atmosphere	climate

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
HAZARDS Earthquake	earthquake	Geoscientific Information	geological process
HAZARDS Fire	fire	Health	disaster
HAZARDS Flood	flood	Inland Waters	hydrosphere
HAZARDS Flood	flood	Oceans	hydrosphere
HAZARDS Landslip	landslide	Environment	land
HAZARDS Manmade			
HAZARDS Pests	pest	Biota	biosphere
HAZARDS Severe local storms	weather	Climatology/Meteorology/Atmosphere	climate
HAZARDS Tsunamis	seismic sea wave	Inland Waters	hydrosphere
HAZARDS Tsunamis	seismic sea wave	Oceans	hydrosphere
HEALTH	health	Health	health
HERITAGE	cultural heritage	Society	culture (society)
HERITAGE Aboriginal			
HERITAGE Architectural	traditional architecture	Geoscientific Information	science
HERITAGE Natural	natural heritage	Society	culture (society)
HERITAGE World	world heritage site	Environment	land
HUMAN ENVIRONMENT	human habitat	Planning/Cadastre	land setup
HUMAN ENVIRONMENT Economics	economics	Geoscientific Information	science
HUMAN ENVIRONMENT Economics	economics	Geoscientific Information	science
HUMAN ENVIRONMENT Economics	economics	Geoscientific Information	science
HUMAN ENVIRONMENT Housing	housing	Society	society
HUMAN ENVIRONMENT Livability	living environment	Society	society
HUMAN ENVIRONMENT Planning	planning	Planning/Cadastre	planning

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
HUMAN ENVIRONMENT Structures and Facilities	built structure	Structure	built environment
HUMAN ENVIRONMENT Urban Design	urban design	Economy	industrial process
HUMAN ENVIRONMENT Urban Design	urban design	Environment	environmental planning
HUMAN ENVIRONMENT Urban Design	urban design	Environment	environmental planning
INDUSTRY	industry	Economy	industry
INDUSTRY Manufacturing	industrial manufacturing	Economy	firm
INDUSTRY Mining	mining	Economy	resource utilisation
INDUSTRY Other	industry	Economy	industry
INDUSTRY Primary	primary sector	Economy	economy
INDUSTRY Service	tertiary sector	Economy	economy
LAND	land	Environment	land
LAND Cadastre	land register		administration
LAND Cover	land cover	Biota	biosphere
LAND Cover	land cover	Geoscientific Information	environmental data
LAND Cover	land cover	Geoscientific Information	science
LAND Geodesy	geodesy	Geoscientific Information	science
LAND Geography	geography	Geoscientific Information	science
LAND Ownership	land and property register		administration
LAND Topography	relief (land)	Environment	land
LAND Use	land use	Economy	resource utilisation
LAND Valuation	land value	Economy	economy
MARINE	marine environment	Environment	environment
MARINE Biology	marine biology	Geoscientific Information	science

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
MARINE Coasts	coast	Environment	land
MARINE Estuaries	estuary	Inland Waters	hydrosphere
MARINE Estuaries	estuary	Inland Waters	hydrosphere
MARINE Estuaries	estuary	Oceans	hydrosphere
MARINE Estuaries	estuary	Oceans	hydrosphere
MARINE Geology and Geophysics	marine geology	Geoscientific Information	science
MARINE Human Impacts	marine monitoring	Geoscientific Information	monitoring
MARINE Meteorology			
MARINE Reefs	reef	Environment	land
MINERALS	mineral	Geoscientific Information	lithosphere
MOLECULAR BIOLOGY	molecular biology	Geoscientific Information	science
MOLECULAR BIOLOGY Genetics	genetics	Geoscientific Information	science
OCEANOGRAPHY	oceanography	Geoscientific Information	science
OCEANOGRAPHY Chemical	chemical oceanography	Geoscientific Information	science
OCEANOGRAPHY Physical	physical oceanography	Geoscientific Information	science
PHOTOGRAPHY AND IMAGERY	photography	Economy	industrial process
PHOTOGRAPHY AND IMAGERY Aerial	aerial photography	Economy	industrial process
PHOTOGRAPHY AND IMAGERY Remote Sensing	remote sensing	Geoscientific Information	methodology
PHOTOGRAPHY AND IMAGERY Satellite	satellite image	Society	information
POLLUTION	pollution	Environment	pollution
POLLUTION Air	air pollution	Environment	pollution
POLLUTION Noise	noise pollution	Environment	pollution

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
POLLUTION Soil	soil pollution	Environment	pollution
POLLUTION Water	water pollution	Environment	pollution
SOIL	soil	Geoscientific Information	pedosphere
SOIL Biology	soil biology	Geoscientific Information	science
SOIL Chemistry	soil chemistry	Geoscientific Information	science
SOIL Erosion	soil erosion	Geoscientific Information	pedosphere
SOIL Physics	soil science	Geoscientific Information	science
TRANSPORTATION	transportation	Transportation	transportation
TRANSPORTATION Air	air transportation	Transportation	transportation
TRANSPORTATION Land	land transportation	Transportation	transportation
TRANSPORTATION Marine	maritime transport	Transportation	transportation
UTILITIES	public utility	Economy	services
VEGETATION	vegetation	Biota	biosphere
VEGETATION	vegetation	Geoscientific Information	science
VEGETATION Floristic			
VEGETATION Structural			
WASTE	waste	Environment	waste
WASTE Greenhouse gas	greenhouse gas	Environment	waste
WASTE Heat	waste heat	Environment	pollutant
WASTE Liquid	liquid waste	Environment	waste
WASTE Sewage	sewage	Environment	waste
WASTE Solid	solid waste	Environment	waste
WASTE Toxic	toxic waste	Environment	waste
WATER	water (substance)		material
WATER Groundwater	groundwater	Inland Waters	hydrosphere
WATER Groundwater	groundwater	Oceans	hydrosphere
WATER Hydrochemistry			
WATER Hydrology	hydrology	Geoscientific Information	science
WATER Lakes	lake	Inland Waters	hydrosphere

NSW Natural Resources Data Framework Thesaurus Development

sourceTerm	GEMETTerm	topicCategory	topTerm
WATER Lakes	lake	Oceans	hydrosphere
WATER Quality	water quality	Environment	environmental assessment
WATER Rivers	river	Inland Waters	hydrosphere
WATER Rivers	river	Inland Waters	hydrosphere
WATER Rivers	river	Oceans	hydrosphere
WATER Rivers	river	Oceans	hydrosphere
WATER Salinity	water salinity		chemical property
WATER Supply	water supply	Economy	services
WATER Surface	surface water	Inland Waters	hydrosphere
WATER Surface	surface water	Oceans	hydrosphere
WATER Wetlands	wetland	Environment	land

6 MAINTENANCE OF THE THESAURUS

The aim of the maintenance program for the thesaurus is to continue in the development of an agreed taxonomy and classification of natural resource terms. Both of these are subject to change and consensus agreement is required before changing the thesaurus. In particular care needs to be taken with any change to the classification as previously indexed data sets may then be incorrectly classified.

Listed below is the proposed maintenance sequence of tasks to bring the current first draft to a level where it could be published and referenced.

1. Publish the draft thesaurus

Issue the thesaurus as text listings and Word documents to all NRIMS agencies.

2. Agency review of the thesaurus philosophy

The review should firstly ratify the philosophy behind the development of the draft thesaurus. This philosophy can be summarised as:

- Use the ISO 10155 TopicCategory terms as the highest level classification term in the thesaurus. These terms were chosen as they will become mandatory reporting requirements for data set listing under this standard and under the ANZLIC profile of this standard.
- Adopt the GEMET terms and their hierarchy. This means using the GEMET terms as developed and agreed in Europe and the USA.
- Extend the GEMET terms for those areas not adequately covered for Australian natural resource and environment use. For instance there does not exist any terms related to Native vegetation, indigenous issues, remediation etc. These should be added to the terms list and appropriate relationships (broader terms/narrower terms) developed.

3. Agency review of the thesaurus content

This review will involve reviewing those areas of the thesaurus for which the Agency has subject matter experts. The review should include the review of the terms lists candidates and particularly their relationships. The appendices to this report lists candidate terms taken from many agencies. Each term in each list should be reviewed.

- If the term exists in GEMET then no action need be taken
- If the term does not exist but an alternate term has been selected (synonym) then that synonym should be verified.
- If no term exists then that term should be promoted for consideration as an addition to the thesaurus.

In each case the hierarchy of the terms needs to be reviewed as well. In the case where no term exists this hierarchy must be established to determine where the new term is added. All revisions are to be presented to the Thesaurus administrator.

4. Establish a Thesaurus administrator.

This officer will be responsible for:

- Obtaining and operating software that will allow the term relationships in the thesaurus to be maintained (this is discussed in a later section).
- Publishing the thesaurus in form suitable for agency use/adoption
- Assisting Agencies to index data sets
- Managing the process of promoting new terms to the thesaurus
- Managing the versioning of the thesaurus and archiving of all previous published versions
- Managing the change control procedures for the thesaurus. This would include seeking approval from a Steering Committee if a term relationship changes.
- Managing a data environment consisting of:
 - Current production version of the thesaurus
 - Development version of the thesaurus
 - Next release version of the thesaurus
 - All previous production versions as an archive.
- Reporting to the Thesaurus Steering Committee

5. Establish a Thesaurus Steering Committee

A Steering Committee needs to be established to approve all changes to the published thesaurus. The Steering Committee representation would comprise all NRIMS Agencies along with representative for OIT, State Library and the CANRI web site maintainer.

The Steering Committee secretary would be the Thesaurus administrator.

The Steering Committee would also be responsible for promoting any changes to the TopicCategory terms as proposed. This would need to be promoted to the ISO for further review.

7 REALISING BENEFITS FROM THE THESAURUS

As the developer of the draft natural resources and environment thesaurus NRIMS can realise the following benefits from the thesaurus:

1. Publication of the agreed thesaurus (following Agency revision and approval) will give the NSW Natural Resources and Environment Agencies a comprehensive indexing tool for indexing natural resource and environmental data. The thesaurus will allow a much finer term to be applied to data sets than is currently possible with the ANZLIC subject keywords.
2. Provide a variety of search mechanisms for natural resource and environmental data. The adoption of the thesaurus allows three methods of searching for data sets on the web to be presented:
 - In the first method the user is asked for a search term and this is used. This method provides the user with no assistance in selecting a term and assumes they are familiar with the terminology.
 - The second method provides the user with access to the thesaurus hierarchy. In this case the user is presented with the TopicCategories as the first level of searching:

Biota
Climatology/Meteorology/Atmosphere
Economy
Environment
Farming
Geoscientific Information
Health
Imagery/Base Maps/Earth Cover
Inland Waters
Planning/Cadastre
Society
Structure
Transportation
Utilities/Communication

Once one is selected, “Geoscientific Information” for example, then all the top terms available under that term are displayed:

Analysis
Environmental data
Evaluation
Experiment
Geological process
Lithosphere
Methodology
Monitoring

Parameter
 Pedosphere
 Research
 Science
 Statistics

Selecting “pedosphere” then displays two further terms:

Soil
 Soil process

Selecting “soil” then displays eight more terms:

Soil air
 Soil function
 Soil profile
 Soil salinity
 Soil structure
 Soil texture
 Soil type
 Soil water

Selecting “soil profile” displays the data set in question.

At each stage in the process all the data sets that have the index words selected could be displayed. Thus under the first selection “Geoscientific Information” there could be thousands, under the next term “Pedosphere” hundreds, and so on.

- The final method of searching involves a pre-defined template of terms from the thesaurus. The terms used in the template do not need to be in their correct hierarchy and do not need to include all terms. For instance the top level of terms presented could list say only 10 of the 19 TopicCategories. In this way data sets on “Health”, “Society”, “Utilities” etc could be removed from the view of the data sets that is being presented. Similarly if a level below is shown in the template it too could select only those terms that the web administrator wishes to provide access to.

For instance a predominantly environmental web site could present a structure as follows:

Biota		
Biochemical process		
Environment		
Environment	assessment	control
impact	management	Planning
Policy	Problem solving	Protection
Pollution and Waste		
Pollution	Pollutants	Waste

Each of these terms would display all data sets that included these index terms and those below them in the hierarchy.

3. Encourage Agency custodianship over logical parts of the classification. Currently data custodianship is based at the data sets level. The development of a logical and consistent classification of natural resource and environmental terms can be used to develop similar custodial responsibilities. Custodianship probably lies at the Top Term level of the classification . This extension to the current concept of custodianship should be further investigated and can be initiated by mapping Agency programs under this classification to determine who is active in each area. The data set classification which is contained in Appendix 3, contains a comprehensive list of data sets classified according to their Top Term. This list could be used as a starting point to map Agency custodianship over these terms. The results are preliminary in that the mapping has been made using the Agency defined ANZLIC subject keywords. With the availability of more index words through the thesaurus the top terms assigned to Agency data sets may change.
4. NRIMS can gain a considerable degree of consistency across Agency web sites by encouraging the adoption of the thesaurus and its classification. Individual Agencies need only to refer to those elements of the thesaurus they wish to and can present these on their web sites as a simple classification pointing directly to data sets they wish to publish on the web.

8 THESAURUS SOFTWARE REVIEW

8.1 Web references for thesaurus software

Searches were made for maintenance software and reference papers over the Web and recommended references are summarised below.

8.1.1 Australian Society of Indexers

<http://www.aussi.org/resources/software/thesauri.htm>

Australian Society of Indexers web site recommends Michael Middleton's web site at the Queensland University of Technology for references to thesaurus management software and related material.

8.1.2 Michael Middleton (Queensland University of Technology)

http://www2.fit.qut.edu.au/InfoSys/middle/cont_voc.html

Michael references:

- American Society of Indexers thesaurus management software
www.asindexing.org/site/thessoft.shtml
- Term Tree 2000
www.termtree.com.au
- THESmain for mono- and multilingual thesauri
www.cedar.at/wgr_home
- WebChoir
www.webchoir.com
- Willpower Information software for building and editing thesauri
www.willpower.demon.co.uk/thessoft.htm

8.1.3 American Society of Indexers

<http://www.asindexing.org/site/thessoft.shtml>

The society web site lists the following thesaurus maintenance packages:

- Data Harmony
www.dataharmony.com
- LEXICO
www.pmei.com/lexico/lexico.html
- MultiTes
www.multites.com
- STRIDE
www.questans.co.uk/p10012.html
- Synaptica
www.synaptica.com/asis99.asp
- TCS-8 and TermChoir
www.webchoir.com/TCS-8.html and www.webchoir.com/TermChoir.html

- Term Tree 2000
www.termtree.com.au

8.1.4 Willpower

<http://www.willpower.demon.co.uk/thessoft.htm>

Willpower Information provides several papers on thesaurus software at their web site. These are:

Ganzmann, J., 1990

“Criteria for the evaluation of thesaurus software” International Classification 17(1990) No. 3/4, pages 148-157, International Society for Knowledge Organisation (ISKO), Ergon Verlag, 1990

<http://www.willpower.demon.co.uk/ganzmann.htm>

Will, L., 2002

“Software for building and editing thesauri”, June 3, 2002

<http://www.willpower.demon.co.uk/thessoft.htm>

The latter paper provides a comprehensive list of software for building and supporting thesaurus.

8.1.5 State Recordkeeping, New South Wales

<http://www.records.nsw.gov.au/publicsector/rk/rib/mergeaaa.htm>

The site provides an overview of the steps involved in compiling a functional thesaurus and merging it with Keyword AAA. Keyword AAA, produced by the Archives Authority of NSW, is a thesaurus of general terminology designed for use in classifying, titling and indexing administrative records in most technological environments.

8.2 Thesaurus software available

A list of suitable thesaurus maintenance software is listed here which are believed to be capable of providing the required maintenance.

The software includes

- Term Tree 2000
- MultiTes
- WebChoir
- Oracle Text (formerly *interMedia* Text)

Summary notes on this software are given in the later sections.

8.3 Thesaurus software requirements

The proposed thesaurus is built up from ISO 19115 Topic Categories as the highest level terms and these are mapped to the top terms (top level) of GEMET (Generalised Multilingual Environment Thesaurus).

Using a maintenance software product such as TermTree.

The 19 ISO Topic Categories would be entered as a level 1 term. GEMET top terms would then be entered at level 2 and linked to the topic categories. The level 2 terms would then be linked to the level 1 terms.

The management software enables terms to be added and linked to the software using drag and drop techniques in an environment similar to Windows Explorer. However trees cannot be dragged.

The GEMET thesaurus terms could be imported through the file formats provided. The software suppliers also provide a service to build the thesaurus.

8.4 Term Tree 2000

<http://www.termtree.com.au>

Term Tree 2000 Editor Standalone Editor

Read/Write administrator module to Create, Edit and Navigate custom thesauri. A full function, limited usage, version of this product is provided with the downloadable sample and on the demonstration CD.

Term Tree 2000 Navigator Standalone read only Navigator

Allows a user on a standalone or Intranet system machine to access the information in one or more locally or Intranet connected thesauri. A full function, limited usage, version of this product is provided with the downloadable sample and on the demonstration CD.

8.4.1 Supplier

This to That Pty Ltd
216 Mount St
Fernmount NSW 2454
Australia

termtree@tpg.com.au
(02) 6655 9059

8.4.2 Notes

Databases supported are:

- Access 97
- Access 2000
- SQL Server 7
- Oracle 8

8.4.3 Import formats

Tag file import in

- “Trim Records Management Thesaurus” format
- “Objective Record Management Keyword Export”
- import Term Tree file structure.

This import is an ASCII text format, one definition field per line. Each line is tagged to indicate the data field.

8.4.4 Cost

Single user standalone Editor/Navigator kits start from \$A880.00 (inclusive of GST).

Special Educational use licences are available.

Multi-User and Corporate installation prices are available.

8.5 MultiTes

<http://www.multites.com>

MultiTes makes it easy to create, browse, print, and maintain different types of vocabularies and thesauri on a standalone PC or a Local/Wide Area Network. MultiTes allows you to handle lists of Terms, Descriptors, Top Terms, Subject Headings, Taxonomies, Polyhierarchical Thesauri, Authority Files, Subject Categories, and Multilingual Thesauri.

MultiTes features:

- Up to 100 million terms per thesaurus
- Unlimited number of thesauri
- Unlimited number of hierarchies per thesaurus
- Unlimited number of relationships for each term
- Support for ANSI/NISO standard relationships
- Support for user-defined relationships and comment fields
- Validation of conflicting relationships
- Automatic generation of reciprocal relationships
- Support for both monolingual and multilingual thesauri
- Support for foreign characters (diacritical marks)
- Quick search of keywords and categories
- Support for multiple users (LAN)
- Support for multiple categories/classification (Microthesauri)
- Support for polyhierarchical thesauri (multiple BT's)
- Reports: Alphabetical, TopTerm, Hierarchical, Classified, Rotated, HTML

Modules include:

- Import Module: Allows you to import terms, relationships, or fully-featured thesauri from ASCII (text) files
- Export Module: Allows you to export your reports to text files in order to interact with other software packages
- Report Generator: Allows you to define your own reports
- HTML File Generator: Creates all the HTML files necessary to publish your thesaurus on the Internet

8.5.1 Supplier

Multisystems
P.O. Box 833205
Miami, FL 33283

Toll Free in the U.S. and Canada: 1-888-45-Multi (1-888-456-8584)

From other countries: (786) 242-1313

Fax: (786) 242-5534

E-mail: MultiTes@aol.com

8.5.2 Notes

Languages

Can support up to 12 languages.

Systems supported

- Windows 95/98
- Windows NT
- Windows ME
- Windows 2000

8.5.3 Import formats

The Import Module can be used to import a thesaurus (or any list of terms) from text files. The import format is a tagged ASCII.

8.5.3.1 ANSI/NISO Z39.19 guidelines

MultiTes can build ANSI/NISO standard thesauri. Each thesaurus is created with the following standard relationships:

USE: Use Preferred Term

UF: Used for

BT: Broader Term

NT: Narrower Term

RT: Related Term

SN: Scope Note

User-defined relationships or comment fields can be added.

8.5.4 Cost

Single-user (upgrade): \$125.00 plus shipping and handling

Single-user Full License: \$295.00 plus shipping and handling

Multi-user LAN License

Site License (unlimited users per location)

Institutional License (unlimited users worldwide)

Educational/Government Agencies

8.6 WebChoir

<http://www.webchoir.com>

8.6.1 Software available

WebChoir products offer solutions for developing, managing, and publishing finding aids and indexing languages such as thesauri, taxonomies, and classifications -- inhouse and for your clientele, without having to modify for one or the other.

TermChoir

A revolutionary hierarchical information organizing and searching tool that enables information professionals to create and search upon varieties of hierarchical subject categories, controlled vocabularies, and taxonomies based on either pre-defined standards or a user-defined structure. It is open, scalable, secure, flexible, and multi-lingual. It supports most popular platforms (NT or UNIX) and databases (Oracle or SQL or Sybase). Users access TermChoir either via client/server connectivity or through internet/intranet/extranet web browser access.

Moreover, TermChoir allows thesaurus producers to package their intellectual property for ultimate customization by unique sets of users. This allows such clients as information architects and interface designers, to name a few, to customize category labels originating from licensed vocabularies to meet internal file designs, or to meet the requirements for unique search/browse schemas for their own products.

ViewChoir

ViewChoir, a read-only version of TermChoir, facilitates the electronic publishing of thesauri, allowing for multiple avenues for commercial distribution of intellectual property across the globe. ViewChoir is an ideal way for thesaurus producers to generate ready reference for staff developing vocabularies, or for clients in the field who may constitute an established subscription base in conjunction with other products or resources. ViewChoir is excellent for constructing collaborative environments within your organization or with cooperating user groups, enhancing the effectiveness and reach of your controlled vocabularies and the products and resources that they supplement.

LinkChoir

LinkChoir is the key to automatically and efficiently exploiting thesaurus products and resources in a networked environment. LinkChoir enables thesaurus producers to consolidate thesauri and indexing vocabularies with a diverse range of client information sources. Indexers, knowledge engineers, digital asset managers, and information seekers alike increase productivity significantly by automatically accessing thesauri that are co-joined via LinkChoir to the records being searched or indexed. LinkChoir eliminates the usual monotonous transfer steps taken when applying terms from vocabulary resources during indexing and searching, saving time and money. Information professionals can use LinkChoir to browse or search controlled vocabularies and assign the appropriate descriptors to the records in their databases. LinkChoir not only allows the indexer to describe information sources using terminology organized in TermChoir, but also provides them with a utility to update the descriptors that they have used to index their information sources.

SeekChoir

Since building the topic hierarchy is the single largest expenditure of initial corporate portal development, information architects and knowledge engineers seek standard thesauri, taxonomies, and classifications. SeekChoir has been tailored to aid thesaurus producers in meeting this demand. SeekChoir liberates the producer's valuable intellectual property by integrating and consolidating thesaurus content with advanced search mechanisms. Publishers easily import their thesauri into the SeekChoir product for distribution, and then their clients have a complete browsable and searchable concept hierarchy for a diverse range of search, data mining, and knowledge management needs.

SeekChoir is a complete, full-featured retrieval system that enables users to browse thesaurus descriptors, allowing the searcher many ways to investigate and employ related and synonymous topics and concepts while searching. SeekChoir also allows seekers to search databases indexed by thesauri constructed in TermChoir or some other thesaurus builder application. SeekChoir supports vocabulary-enhanced keyword searches within database records. But, if LinkChoir has been employed to index database records, SeekChoir also supports a controlled vocabulary search of those records indexed with terms from the associated thesaurus.

SeekChoir is expected to be available by the end of March 2002.

8.6.2 Supplier

Sales & Marketing
11301 Olympic Blvd.
Suite 586
Los Angeles, CA 90064

Phone: (310) 966-1558
Fax: (310) 966-1558
sales@webchoir.com

8.6.3 Notes

Systems

- Windows - Microsoft Windows NT/2000 Server
- UNIX
- Macintosh

8.6.4 Cost

TCS-8 Licensing Fees

Single User: \$2,795 (plus \$200 for Borland Interbase) Multi-User: \$3,995 (plus \$200 each per Borland Interbase user, maximum 5)

Single User competitive upgrade: \$ 995 (plus \$200 for Borland Interbase)

Multi-User competitive upgrade: \$1995 (plus \$200 each per Interbase user, maximum 5)

Single User TCS upgrade: \$ 795 (plus \$200 for Borland Interbase)

Multi-User TCS upgrade: \$1795 (plus \$200 each per Interbase user, maximum 5)

8.7 Oracle Text (formerly *interMedia Text*)

<http://technet.oracle.com/products/text/content.html>

8.7.1 Software available

Oracle Text uses standard SQL to index, search, and analyse text and documents stored in the Oracle database, files and on the Web.

Oracle Text can analyse document themes and gists;

- search text using variety of strategies including full-text: boolean, exact phrase, proximity, section searching, misspellings, stemming, wildcard, thesarus, word equivalence, and scoring, mixed: text index plus relational attributes, and thematic;
- search HTML and XML sections and tag values;
- render search results in various formats including unformatted text, HTML with automatic keyword highlighting, and original document format;
- analyze and index most document formats with over 150 document filters;
- supports 39 languages;
- bulk load documents in Oracle8i/9i with SQL*Loader.

8.7.2 Supplier

Oracle

8.7.3 Cost

Believed included with Oracle 9.

9 DRAFT THESAURUS DATABASE

9.1 MS Access database

The Thesaurus has been developed in MS Access. This was the original form that the GEMET thesaurus was supplied in. This database has been modified to remove all, the foreign language terms to reduce the size of the database (from 18 M byte to 2 M byte). In addition the GEMET Groups and Themes tables have been removed as these have been replaced.

9.2 Terms database

9.2.1 Database tables

The terms database called NaturalResourcesDataFrameworkThesaurus_v14 is in Microsoft Access 97 and contains the following tables:

Table names	Descriptions
GEMET	
Terms	List of terms
NonPrefTerms	List of non preferred terms
Terms_to_BroadNarrowTerms	pointers between broad and narrow terms
Terms_to_Non_Pref_Terms	pointers between terms and non preferred terms
ISO topic categories	
TopicCategories	ISO topic categories
TopicCategories_to_Topterms	ISO topic categories mapped to GEMET topterms
NaturalResourcesDataFrameworkThesaurusTerms	
Natural Resource Datasets	
Datasets	Natural resource dataset names provided by CANRI
DatasetSearchWords	Dataset search words for the above datasets
DatasetClassification	Datasets with associated hierarchy of GEMET terms
Candidate classifications	
TermsANRA	ANRA
TermsANZECC	ANZECC Core Environmental Indicators
TermsCatchmentBlueprint	Catchment blueprint
TermsEPASoE2003	EPA State of Environment chapter headings
TermsGSDI	Global Spatial Data Infrastructure
TermsLocalGov	Local Government
TermsManly	Manly Environment Centre terms
TermsNRIMS	NRIMS Strategy Themes
TermsNRMTheme	NRM Themes

NSW Natural Resources Data Framework Thesaurus Development

Table names	Descriptions
Candidate terms	
TermsANZLIC	ANZLIC search words and qualifiers
TermsCERES	CERES terms
TermsDLWC	Department of Land and Water Conservation terms
TermsEA	Environment Australia
TermsEPASoEWeb	EPA State of Environment Web site terms
TermsNAL	NAL Agricultural Thesaurus terms
TermsNSWEPA	NSW EPA thesaurus
TermsNTSupSci	NT Supervising Scientists terms
TermsUSLibCongress	US Library of Congress terms
TermsWISE	DLWC Water Information System
Candidate terms with hierarchy	
TermsANZLICTrees	Terms with associated hierarchy of GEMET terms
TermsCERESTrees	
TermsDLWCTrees	
TermsEATrees	
TermsEPASoEWebTrees	
TermsManlyTrees	
TermsNALTrees	
TermsNSWEPATrees	
TermsNTSupSciTrees	
TermsUSLibCongressTrees	
TermsWISETrees	

9.2.2 Attributes

The attributes for these tables are given in the tables below:

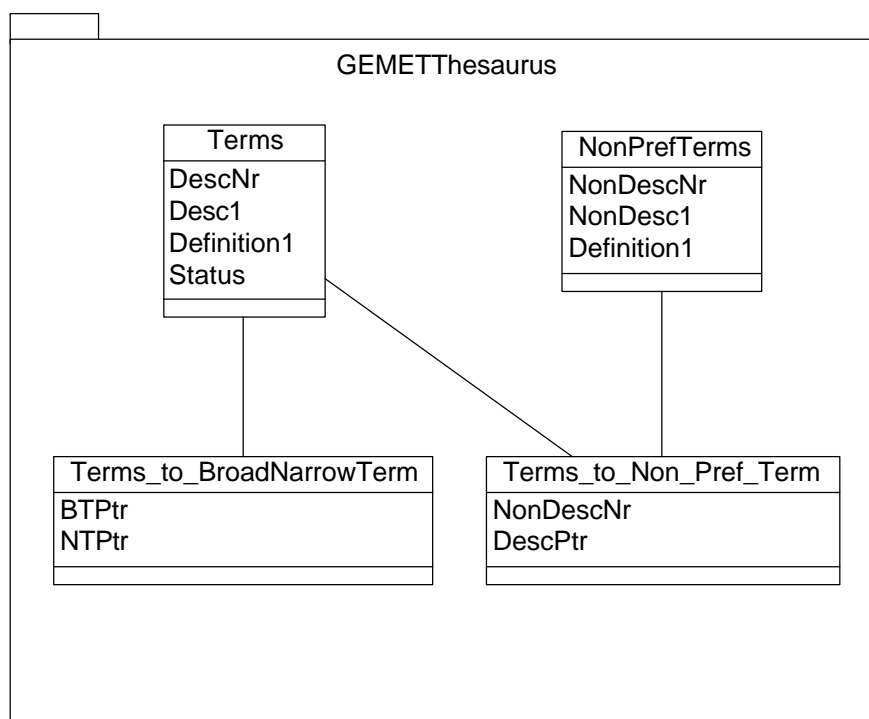
The attribute and table names are written in the UML data modelling conventions except where the tables have been imported from other sources.

Attributes	Descriptions
GEMET tables	
Terms	List of terms
DescNr	term number (identifier)
Desc1	term
Definition1	term definition
Status	status of term (topterm, relations, synonyms, unused)
.....	
NonPrefTerms	List of non preferred terms
NonDescNr	non preferred term number (identifier)
NonDesc1	non preferred term
Definition1	term definition
.....	
Terms_to_BroadNarrowTerms	pointers between broad and narrow terms
BTPtr	Broad term pointer from term table
NTPtr	Narrow term pointer from term table
.....	
Terms_to_Non_Pref_Terms	pointers between terms and non preferred terms
NonDescNr	no preferred term number
DescPtr	term number pointer
ISO topic categories	
TopicCategories	ISO topic categories
topicCategoryId	topic category identifier (text)
topicCategory	topic category name
TopicCategories_to_Topterms	ISO topic categories mapped to GEMET topterms
topicCategoryId	topic category identifier (text)
topicCategory	topic category name
termId	GEMET topterm number and key
term	GEMET topterm
Candidate terms	
TermsANZLIC (for example)	ANZLIC search words and qualifiers
term	search words and qualifiers

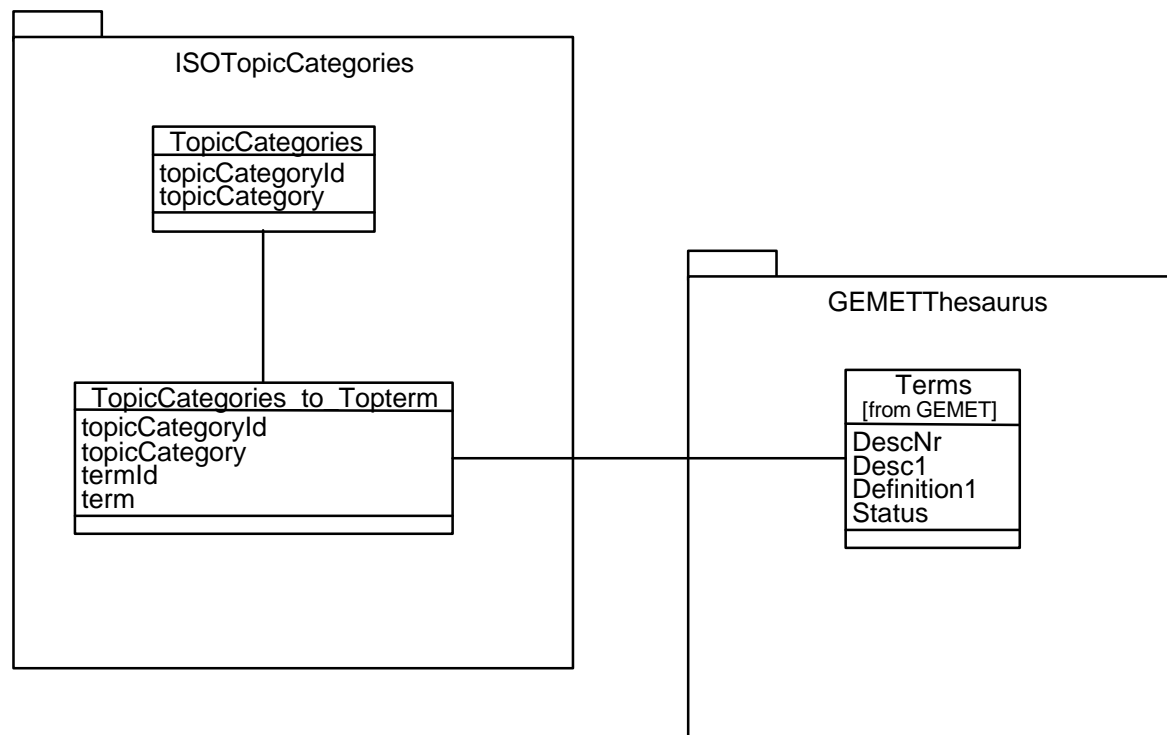
Attributes	Descriptions
searchWord	search word
qualifier	qualifier
source	source: ANZLIC
alternateTerm	mapped GEMET term where required
type	type of GEMET term: term or non preferred
GEMETTermId	GEMET term identifier
Candidate terms with hierarchy	
TermsANZLICTrees (for example)	Terms with associated hierarchy of GEMET terms
sourceTerm	ANZLIC search word
alternateTerm	mapped GEMET term where required
GEMETTerm	GEMET term corresponding to search word
topicCategory	topic category at top of term hierarchy
topTerm	top GEMET term below topic category
level1	level 1 term
.....	down to term corresponding to search word
level10	level 10 term if required

9.2.3 Data models

The GEMET data model relationships are summarised in the diagram below.



The ISO topic categories are related to the GEMET Thesaurus terms in the following diagram



10 REFERENCES

1. Description of NSW Government Information and Services using AGLS Subject. Consultant's Report. State Library of New South Wales. 29 February 2000
2. The World of multilingual thesauri. Jacek Purat. School of Information Management and Systems. University of California, Berkeley 1998.
http://www.sims.berkeley.edu/~purat/world_multilingual_environmental_thesauri.html
3. Core environmental indicators for reporting on the state of the environment. Australia New Zealand Environment and Conservation Council State of the Environment Reporting Task Force. March 2000
4. ANZLIC Metadata Guidelines: Core metadata elements for geographic data in Australia and New Zealand. Version 2 February 2001. Australia New Zealand Land Information Council – Metadata working group.
5. Community Access to Natural Resource Information CANRI website.
www.canri.nsw.gov.au
6. Living Thing web Site. Service NSW. www.livingthing.net.au
7. CERES. California Resources Agency. <http://ceres.ca.gov>
8. DLWC Thesaurus (Draft version 0.1, January 2002)
9. Environment Australia. ea-thesaurus.doc (as supplied)
10. NSW EPA web site www.epa.nsw.gov.au
11. Global Spatial Data Infrastructure. <http://130.11.52.184/servlet/FGDCWizard>
12. Environmental Guidelines. State of the Environment Reporting by Local Government December 1999. www.dlg.nsw.gov.au
13. Judy Reizes, pers comm. Manly Environmental Centre
14. NAL Agricultural Thesaurus. <http://agclass.nal.usda.gov/agt/agt.htm>
15. Natural Resource Management Ministerial Committee Framework. Minimum set of indicators and protocols for national review. Supplied by Paul Elton DLWC.
16. National Land and Water Resources Audit. Australian National Resources Atlas.
http://audit.ea.gov.au/ANRA/atlas_home.cfm
17. Template for regional catchment plans/blueprints. Mike Thompson Nature.net
18. . NSW State of the Environment 2003 Chapter Headings (Jan02). Provided by Peter Maganov EPA
19. . NSE EPA Thesaurus. Supplied by David Large EPA
20. Northern Territory Office of the Supervising Scientist
21. Sydney Olympic Park Authority Thesaurus. Supplied by John Hudson
22. US Library of Congress Subject Headings. 24th Edition 2001.
23. DLWC WISE system. Provided by Richard Kingsford.

11 APPENDIX 1 – REVIEW OF TERMS LISTS

The candidate terms lists are attached as Microsoft Excel files. Each file lists the candidate terms and the full hierarchy of the corresponding term in the proposed thesaurus. This can be up to 10 levels deep.

The following candidate terms lists are presented:

Source of Candidate Terms	Description	Excel File
ANZLIC Metadata Guidelines	Subject keywords	TermsANZLICTrees
CERES	Californian environmental thesaurus aligned with GEMET	TermsCERESTrees
DLWC Glossaries	Internal information management system	TermsDLWCTrees
Environment Australia Thesaurus		TermsEATrees
EPA SoE Web site terms	taken from web site classification	TermsEPASoEWebTrees
Manly Environmental Centre	Internal classification	TermsManlyTrees
NAL Agricultural Thesaurus (USA)	US Agricultural thesaurus	TermsNALTrees
NSW EPA Thesaurus		TermsNSWEPATrees
NT Office of Supervising Scientist Thesaurus		TermsNTSupSciTrees
SOPA Gemet thesaurus as revised	Sydney Olympic Park Authority thesaurus developed from GEMET and extended for remediation terms	Not included
US Library of Congress	Partial listing for ecology and nature conservation	TermsUSLibCongressTrees
WISE (DLWC)	DLWC Water Information System CD	TermsWISETrees

12 APPENDIX 2 – REVIEW OF CLASSIFICATIONS

See separate volume.

13 APPENDIX 3 – CANRI CORE AND PRIORITY DATA SET CLASSIFICATION

See separate volume.