Enhancing a subject vocabulary for Australian education

Philip Hider, Barbara Spiller, Pru Mitchell, Robert Parkes and Raylee Macaulay

This paper presents a case study focusing on the process of improving subject access to a collection of resources for teachers in the higher education sector. It describes how existing controlled indexing vocabularies in the education field were evaluated as candidates for adoption, according to a set of nine evaluation criteria, including the depth to which the vocabularies cover the subfield of higher education, their linguistic and ontological alignment with the target audience, their levels of maintenance and institutional support, and the extent to which they conform to international standards.

The paper then describes how the Australian Thesaurus of Education Descriptors (ATED) was selected as the closest to the requirements, but that as a general education vocabulary it lacked some of the specificity needed to describe higher education resources. Areas where ATED’s coverage was inadequate were identified via a mapping exercise. Selected terms were then assessed for their suitability to be incorporated into ATED.

This paper highlights how established vocabularies can not only be used in different environments, but can themselves be strengthened through the process.
Background

The Australian Government Office for Learning and Teaching (OLT) has provided considerable amounts of funding ($57.1 million in 2012-15; OLT, 2014) to projects investigating higher education learning and teaching practices in Australia. Its online Resource Library (www.olt.gov.au/resource-library) is a key means of disseminating the outcomes of these projects.

Presently, in 2015, the Resource Library houses resources from over 600 projects, funded not only by the OLT, but also by its predecessor organisations, including the Australian Learning and Teaching Council (ALTC, 2009-2011) and the Carrick Institute for Learning and Teaching in Higher Education (2004-2009).

In order to find these resources and apply the recommendations emanating from these projects, it is vital that educators and other stakeholders are aware of, and effectively able to use, the Resource Library. However, anecdotal evidence indicated a lack of awareness of the Resource Library amongst Australian academics and problems with consistently being able to find and retrieve relevant resources from the database. This situation prompted the OLT to commission a project in 2014 to redesign the Resource Library, in order to improve access and usability.

The project, conducted by a team of academics and librarians from Charles Sturt University, the University of Wollongong and the Australian Council for Educational Research (ACER), was completed in 2015. It aimed to deliver a new, re-indexed database for the OLT, which provides improved access to the content of the Resource Library for end users.

The user survey conducted in the first phase of this project pointed to the usefulness of subject searching (Hider et al., 2015, p.8). Almost all respondents (97%) rated topic/subject as a useful field by which to search the Resource Library (author was rated the second most useful at 74%). However, almost 1,600 keywords had been used to describe the subjects of the project resources, and it was clear that they had been used very inconsistently and without reference to the collection as a whole. The project team determined that subject access would be improved by replacing the current keywords with descriptors derived from a controlled vocabulary.

Selection of an appropriate controlled vocabulary involved development of a set of evaluation criteria, identification of candidate vocabularies and then a process of evaluation to determine if there was an existing vocabulary that could be used for the purpose. This vocabulary could then be used in the re-indexing of all the existing resources stored in the database, and for the indexing of new resources entered into the database in the future. The process of re-indexing involved the editing of all existing records for the resources, including the replacing of the existing, uncontrolled keywords with descriptors from the new subject vocabulary.

Vocabulary evaluation

Looking for existing criteria on which to base the evaluation of education vocabularies led the project team to concur with Pinto (2008), who found that ‘while much has been written on thesaurus construction and use, few publications explicitly approach thesaurus quality as a single concept.’ Pinto went on to identify and study four factors of thesaurus quality – conceptual framework, performance, format and external aids – which provide a good foundation on which to build a set of criteria. Golub and Tudhope (2008) reviewed vocabulary registries to identify elements useful in facilitating the discovery of a relevant thesaurus, grouping these elements into five categories, namely general information, scope and usage, vocabulary characteristics, terms and conditions, and provider. They suggest that these categories could also be useful for evaluation of the vocabulary for a particular application or use. Owens and Cochrane (2004) considered structural criteria and adherence to vocabulary standards, as well as criteria related to the form of words used in the vocabulary. Liese (2008) focused on issues of governance and maintenance of controlled vocabularies, including the change control process, policy on reviews, and ownership.

In determining the process for evaluating a subject vocabulary, the standard manual for thesaurus construction by Aitchison, Gilchrist and Bawden (2000) also provided guidance, for as Losee (2007) contends ‘criteria that are sufficient for development are also sufficient to measure performance.’ For Lykke-Nielsen (2000), thesaurus construction requires both domain knowledge, which covers knowledge about the environment and the situations in which the thesaurus is to be used, and conceptual knowledge, that is, the semantic and linguistic understanding which informs the structuring of the vocabulary. It also demands attention to four aspects: ‘the system users and their information needs and search behaviour; the subject field; the use of language; and the type of data/literature, its quantity, and the available resources’ (Lykke-Nielsen, 2000). The tool used to host a vocabulary was a consideration raised by Martinez-Gonzalez and Alvite-Diez (2014, p.711), who argued that it is important for SKOS and RDF to be supported, as ‘the two semantic web standards that will permit thesauri to achieve a successful place in information interoperability, which is so important in web information systems.’
In addition to the criteria that apply to all vocabularies, there were considerations specific to the content and scope requirements of this project. The following nine vocabulary evaluation criteria were agreed upon by the project team. The first five criteria related to the content of the vocabulary, that is, its concepts, terminology and structure. The final four criteria considered elements related to the vocabulary carrier, that is, the system used to host the vocabulary and the organisation maintaining it.

**Vocabulary evaluation criteria**

1. **Concepts and terminology**
   To what extent are the concepts and terminology of the vocabulary appropriate to the audience and collection?
   
   This constituted the fundamental issue of conceptual coverage and relevance, as well as the form of words used to express these concepts. For this project the specific audience was Australian learning and teaching practitioners and policy makers in higher education. Five typical concepts used in real OLT search queries were used to compare the vocabularies: ‘blended learning’, ‘curriculum renewal’, ‘work integrated learning’, ‘student retention’ and ‘leadership’.

2. **Structure**
   To what extent is the vocabulary structured consistently, and does it follow vocabulary standards?
   
   The gold standard here would be a thesaurus that expressly complies with the ISO 25964 (2011). It can be difficult to confirm this compliance, but statements by the thesaurus publisher served as an indication. Inclusion of hierarchical relationships, equivalence relationships, associative relationships, scope notes and history notes were elements looked for in this regard. Vocabularies with flat or no obvious structure, and those without elements such as scope notes and qualifiers, received low scores.

3. **Scope**
   To what extent does the breadth of the vocabulary cover the field?
   
   In this project the vocabulary scope was higher education. Vocabularies that focused on higher and vocational education were considered ideal, but broader vocabularies that covered the whole of education, including higher education, could also be considered.

4. **Depth**
   To what extent is the vocabulary precise enough for general users, and reasonably but not overly specific?
   
   To test the depth of each vocabulary the terms ‘degrees’ or ‘levels of education’ were used, and the inclusion of ‘Bachelor’, ‘Masters’, ‘Doctorate’ levels was looked upon favourably. A second test of this criterion was the depth of the concept ‘students’ (were there narrower terms that included types of student?).

5. **Geographic coverage**
   To what extent does the vocabulary reflect the geographic context?
   
   For this project a vocabulary reflecting the Australian context would be ideal, although one with a regional or even global treatment, but with good Australian coverage, could be considered.

6. **Maintenance**
   To what extent is there evidence of ongoing revision and editions of the vocabulary and its terms by an established organisation?
   
   Publication by an established organisation combined with a history of, or strategy for, regular revision provides a certain level of confidence that a vocabulary will continue to be maintained.

7. **Licensing and cost**
   To what extent does use of the vocabulary require payment of upfront or ongoing fees or licence agreement?
   
   The selected vocabulary would ideally be accessible for no charge and downloadable with no licence restrictions. This scenario was rated the highest, with lower scores awarded to vocabularies with licence restrictions or with costs attached to use. It was noted that the costs involved in maintaining a vocabulary could result in a reciprocal relationship between the maintenance and cost criteria.

8. **Usability and support**
To what extent is the user interface and functionality of the vocabulary reliable, accessible and searchable online, and is user support available?

Usability factors such as good search functionality, a welcoming navigable website that invited use of the data, and adherence to website accessibility standards were taken into account. For high scores with respect to user support the project team was looking for full contact details available on the website, prompt responses to emailed enquiries, and straightforward usage and download instructions.

9. Interoperability

To what extent does the vocabulary meet industry standards and requirements of interoperability in terms of its formatting, publication, transmission and encoding?

The importance of selecting a vocabulary that would work well with other systems was paramount given the unknown future of the database and resources which were the focus of this project. Registration of the vocabulary with international standards, such as the MARC Subject Heading source codes (http://www.loc.gov/standards/sourcelist/index.html) and Library of Congress linked data service (http://id.loc.gov), provide some evidence of this. Ideally the vocabulary would be exportable in SKOS format, or, failing that, in XML or as a PDF. Vocabularies with no available export options scored low on this criterion.

Thus the project team had a set of criteria specific to the requirements of the OLT project, but constructed within a framework that should make the criteria relevant to other vocabulary evaluation projects.

Identification of thesauri

The project team then identified candidate thesauri for indexing the Resource Library collection. Potential candidates were collected from vocabularies already known to members of the project team, and from a set of vocabularies found through literature and web searching. The Australian and New Zealand Society of Indexers (ANZSI; 2015) maintains a useful page of vocabularies (http://anzsi.org/resources/reading-lists/thesauri), as does Jisc (2015, http://www.jiscdigitalmedia.ac.uk/guide/controlling-your-language-links-to-metadata-thesauri) and Taxonomy Warehouse (2015, http://www.taxonomywarehouse.com/details.aspx?headwordObject=10061&vunid=89105). For this project, there were two preliminary requirements of the candidate vocabularies. All resources to be described in the project were in English, and this would be the language used by both indexers and users of the resources, so there was a need for the vocabulary to be in English. Also, given the open-ended nature of the indexing, it was considered necessary for the vocabulary to be readily available online. This eliminated a number of commercial vocabularies where access was only available from within a proprietary system.

Expert analysis of candidate vocabularies

Eleven education vocabularies were identified as candidates to be evaluated. Firstly, expert analysis was carried out independently by two indexing specialists from the project team, using the evaluation criteria enumerated in the previous section. For each criterion, the following rating scale was applied.

- 5 = excellent
- 4 = good
- 3 = satisfactory
- 2 = partly unsatisfactory
- 1 = wholly unsatisfactory

Some of the salient features of each of the eleven thesauri (presented in alphabetical order of their abbreviation, as used in this paper) are summarised below.

ATED (Australian Thesaurus of Education Descriptors)


ATED has been produced by the Cunningham Library at Australian Council for Educational Research (ACER) since 1979. It went through four print editions and is now available on the web and updated twice a year. At the time of this project it had 5,033 preferred terms and 4,961 USE references. ATED was well suited to this project’s needs in terms of concepts and terminology, depth, and geographic coverage. Its maintenance, cost, support and interoperability were acceptable. Its weakest point was scope, as its remit is education in general, rather than higher education in particular.
BET (British Education Thesaurus)
http://www.leeds.ac.uk/educol/BEID.html

BET’s focus was on educational policy and administration; evaluation and assessment; technology and special education. Its British bias was evident in terms such as ‘people’s universities’ and ‘sixth form colleges’, but it did include three headings beginning with ‘Australian’ plus ‘Aboriginal Australians’. It had good depth with ‘persons’ NT ‘students’ NT ‘arts students’ NT ‘performing arts students’. BET has been taken over by Ebsco Information Services and it is unclear whether the thesaurus will be updated.

EDU (ÉDUthès: thésaurus de l’éducation)
http://www.cdc.qc.ca/multites_en.htm

ÉDUthès is published by the Centre de Documentation Collegiale in Quebec. In its 6th edition (2001), it included 3,889 terms of which 1855 are preferred terms. It has a Canadian bias, focusing on Quebec in particular (e.g. ‘Canadian history’ NT ‘Quebec history’). It seemed structurally sound, but lacked user-centric terminology, included no mention of Australia, and was too broad (education in general). On the plus side, it was freely available and quite interoperable, with good user support.

EET (European Education Thesaurus)

EET has been published by the Council of Europe since 1969. Its last complete revision was in 2003, but there was evidence of more recent updating (e.g. the term ‘evaluation’ was modified 2007). It had 3,515 terms of which 2,995 were preferred terms. It is European-centred, with names of European organisations and programs. English language headings were presented with translations into eight European languages. Terms beginning with ‘online’ or ‘elearning’ were not found, suggesting a lack of currency.

ERIC (ERIC Thesaurus)
http://eric.ed.gov/?ti=all

ERIC has been published by the US Education Resources Information Center since 1964 and regularly updated, now on a monthly basis. Internationally recognised as an authoritative reference tool, ERIC fulfilled the project’s requirements, but, like ATED, addressed education in general. Its American language (e.g. ‘vacations’, ‘high school freshmen’, and the use of ‘majors’ to describe students rather than programs) was another disadvantage for this Australian project.

EET (European Training Thesaurus)

ETT published by the European Centre for the Development of Vocational Training is in its 3rd edition (2008). ETT’s strengths (for instance terms translated into 12 European languages) did not match well with this project’s requirements. European organisations and programs included were not relevant in an Australian context. It focused on vocational education with some higher education terms, such as ‘university studies’, but not adequate for the purposes of this project.

IBE (UNESCO-IBE Education Thesaurus)

IBE has been published by UNESCO’s International Bureau of Education since 1973. It is in its 6th edition, 2nd revision (2007). The scope is whole of education and international. Its strongest feature for the purposes of this project was its depth, with the ability, for example, to drill down from ‘certification’ through ‘degrees’, to ‘Bachelors’, ‘Doctoral’ and ‘Masters degrees’. Australia was included as a geographic heading, but there were no Aboriginal headings.

SCOT (Schools Online Thesaurus)
http://scot.curriculum.edu.au

SCOT has been published by Education Services Australia since 2010 and was regularly updated. SCOT offers the benefits of an Australian/New Zealand focus (including 19 terms beginning with ‘Aboriginal’), user-centric terminology, and good maintenance, interoperability and support. However, its scope is the school curriculum, and it had less satisfactory coverage of higher education.

TESE (Thesaurus for Education Systems in Europe)
TESE has been published by Eurydice (associated with the European Commission) since 2006. The current edition (2nd) was published in 2009 and has 1,387 descriptors. It focuses on European education systems and policies, including the names of European organisations and programs. Terms are in English, with 14 European language translations. The website has a welcoming and open presentation, providing extensive information about the thesaurus history, revision, structure and objectives. Like ATED and ERIC, its scope is the whole of education.

**UK (‘Education terms’ from the UK Dept of Education)**


This set of UK Education terms has been published by the UK Department of Education since the 1980s. It was last updated in 2013. Not surprisingly it had a British focus, and includes names of British organisations and UK-specific terms (such as ‘student grants’ rather than ‘student loans’). Like most of the vocabularies evaluated, it covered the whole of education. At the time of the evaluation the vocabulary showed strength in its maintenance, availability, user support and interoperability (the ‘container’ criteria). By the end of the project, however, the vocabulary had been archived.

**VOC (VOCED Thesaurus)**

http://hdl.voced.edu.au/10707/6073

VOCED was published in 2006 by the National Centre for Vocational Education Research (NCVER) in Australia. At the time of the evaluation it was no longer used to index NCVER’s own research database, VOCED+, which casts doubt over the future of maintenance of the vocabulary. Like ATED and SCOT it had the benefit of an Australian focus (including six Aboriginal terms), but its scope of vocational education and training was a limitation, and its coverage of higher education terms was inadequate for the purposes of this project. Online access to the vocabulary was removed in mid 2015.

The overall results of the expert analysis are shown in Table 1. ATED scored highest with 81 points out of a possible 90, followed by ERIC and SCOT together on 75. The next highest was TESE on 70. Both experts individually rated ATED highest, scoring it 42 and 39 points respectively.

**Table 1 Expert evaluation scores of vocabularies** here.

The nine criteria were not weighted according to their relative importance, although it should be noted that it would not be desirable to select, for example, a vocabulary that scored strongly on the five ‘carrier’ criteria, if it was weak on the four ‘content’ criteria. The UK government vocabulary, for instance, scored quite highly, but mainly on the basis of carrier criteria.

**Search query matching**

The criteria of concepts and terminology, scope and depth were the focus of the second part of the evaluation process. For this, each vocabulary was tested for its coverage of a list of the most common search query terms logged by the Resource Library system in two one-week periods. For those terms consisting of one word, the following codes were used.

- \( Y \) = match, with the word matching exactly a word amongst the vocabulary’s entry terms or references
- \( N \) = no match, with the word not matching exactly any word amongst the vocabulary’s entry terms or references

For those terms consisting of more than one word, three codes were used, as follows.

- \( F \) = full match, with all the words in the phrase matching exactly words amongst the vocabulary’s entry terms or references; the words may come from different phrases
- \( P \) = partial match, with at least one but not all the words in the phrase matching exactly words amongst the vocabulary’s entry terms or references; the words may come from different phrases
- \( N \) = no match, with none of the words in the phrase matching exactly any of the words amongst the vocabulary’s entry terms or references
The results of the matching exercise are presented in Tables 2 and 3.

**Table 2 Matching of single-word query terms** here.

**Table 3 Matching of multi-word query terms** here.

ATED and SCOT covered the most one-word terms; for the multi-word terms, ATED again provided the most full matches, by some distance, and also the equal top score for most full and partial matches combined.

In summary, the ATED vocabulary presented itself as the best candidate, out of the relevant vocabularies available, for the subject indexing of the OLT Resource Library. Although it was not a perfect fit, covering as it did the field of education more broadly than higher education, it was a very good fit. It should be restated, however, that the evaluation criteria used to select ATED were unique to the needs of this particular project, representing specific scope, geographical and topical requirements.

The project team initially considered implementing the higher education portion of ATED, but the inter-related structure of the thesaurus meant that relevant terms could be found in many facets. Therefore it was decided to adopt ATED as a whole, with a view to extending the thesaurus, where necessary, so that it adequately covered the concepts and terms currently represented in the Resource Library.

**Enhancement of ATED**

Once ATED had been selected as the subject vocabulary of choice for the Resource Library, the project team needed to assess, in more detail, how well the existing thesaurus covered the topics represented in the Library, and which topic areas (if any) needed supplementation. To this end, a mapping exercise was undertaken in which the 1,593 uncontrolled keywords used in the Resource Library were examined for matches with ATED terms. First, exact matches, as well as many near matches that were different only in syntax, were identified. The remaining 1,160 terms were examined by members of the project team from ACER, with input from domain experts. The following questions were considered in the analysis of each keyword:

1. Is the concept represented in ATED? If not, it may be a candidate for inclusion as a new descriptor.
2. If the concept is represented, is the terminology sufficiently different to justify it being added to ATED as a USE reference?
3. If the concept is not represented in ATED, does it have sufficient literary or user warrant to be included?

Answering the third question required consideration of what constituted warrant for this vocabulary. ATED’s primary use is as an in-house indexing tool for ACER’s Cunningham Library staff. It is used to index and search the subject matter of those materials covered by the Australian Education Index (AEI) and its four subsidiary databases, Database of Research on International Education; Education Research Theses; Learning Ground (an Indigenous education research database); and BOLDE, the Blended, Online Learning and Distance Education research bank. When index records are created for AEI, as well as adding subject terms from ATED, natural language terms representing other concepts are added into an ‘identifiers’ field. One way of identifying candidate terms for ATED is to monitor these identifiers for usage trends and upgrade them to ATED descriptors when warranted. Once a threshold of 20 instances of a particular identifier has been reached, consideration is given to its adoption as a new ATED descriptor.

The project team employed a modified version of Cunningham Library’s usual method for assessing new terms for inclusion in its thesaurus. Given the small size of the Resource Library compared to AEI, it would have been difficult for terms to reach the threshold of 20 instances, so in this project, 10 instances was set as the threshold for new descriptors and five for USE references. Usage in AEI was also taken into consideration. However, it was also recognised that usage statistics can never be used alone, as there are other factors to weigh up, such as a term’s currency and its centrality to the field of education.

**Table 4 Categories of terms in OLT-ATED mapping exercise** here.
Table 4 outlines the results of the analysis of existing OLT keywords. The largest grouping was category A. These 934 keywords were considered to be covered both conceptually and terminologically by ATED, although in many cases two or more ATED terms were needed to adequately represent the keyword. For example, ‘academic integrity’ in the Resource Library was covered by the combination of ‘academic conduct’ and ‘integrity’ from ATED, and ‘cross-disciplinary curriculum’ by ‘curriculum development’ and ‘interdisciplinary approach’.

Unlike ATED, which excludes proper nouns, the Resource Library’s keywords included them. The 108 terms in category B were mainly names of organisations, projects and software (e.g. ‘Entry into Valhalla’, a computer program for teaching legal ethics). These were all regarded as out of scope for the purposes of this procedure.

The 65 terms in category C were judged as not being conceptually covered by ATED, but lacked the literary warrant needed to be considered for inclusion. Many of them were either quite specific, or were peripheral to the field of education (e.g. ‘rat dissection’, ‘sonography’ and ‘chest auscultation’).

The two terms in category E were considered to be conceptually but not terminologically covered by ATED, but lacked the literary warrant to be added as USE references. All the terms in categories C and E were nevertheless noted and will be monitored as potential candidate terms for future updates of ATED. In time, sufficient literary warrant may justify their addition.

The 30 terms in category D were judged to be not conceptually covered by ATED, and as having sufficient literary warrant for immediate consideration as new descriptors.

Similarly, the 21 terms in category F, being covered conceptually but not terminologically, had sufficient literary warrant to be considered for immediate inclusion in ATED as USE references.

Dealing with new descriptors

When a new descriptor is being included in ATED, it needs to be incorporated into the existing structural hierarchies, or structural changes need to be made in order to accommodate the term. For each of the new descriptors from the OLT project, a particular structure was proposed, comprising a scope note, broader terms (BT), narrower terms (NT) and/or related terms (RT), along with a rationale for the changes. The proposals were circulated to members of the project team, the project reference group and ACER staff, for comment and discussion. As a result 28 new descriptors and 24 new USE references were ultimately added to ATED.

Discussion

Some aspects of the ATED enhancement process are discussed below.

New concepts for ATED

The exercise illustrated the fact that incorporating a new scope or audience group into an established thesaurus can reveal gaps or bias in the vocabulary. For example, the terms ‘traditional Chinese medicine’, ‘acupuncture’ and ‘naturopathy’ all appeared in the Resource Library keyword list. ‘Alternative medicine’ had been used as an identifier 15 times in AEI, ‘complementary medicine’ five times and ‘Chinese medicine’ twice. This cluster of terms reflects an area of academic research not previously represented in ATED. As a result of this project ‘alternative medicine’ was added as a new descriptor, with USE references ‘alternative therapies’, ‘naturopathy’ and ‘complementary therapies’, and the narrower terms ‘traditional Chinese medicine’, ‘chiropractic’ and ‘acupuncture’. Other new areas of academic study that emerged as worthy of attention were hospitality education, environmental science, environmental engineering, and spatial sciences as an umbrella term for surveying, cartography and geographic information systems.

These new areas gave rise to significant structural questions, however, such as whether ‘alternative medicine’ was a subcategory of ‘medicine’. It was determined that in the field of education, medicine was regarded as a distinct and separate area of study to ‘alternative therapies’, so both were placed under the broader ‘health sciences’.

Adding depth and breadth

The project added depth to ATED in the higher education area, with the addition of terms such as ‘associate degrees’, ‘double degrees’, ‘final year students’ and ‘work integrated learning’. Other new terms, such as ‘distributed leadership’, ‘studios’, ‘teaching research relationship’ and ‘variation theory’, are applicable to all levels of education and thus potentially useful to the field of education more generally. ‘Capabilities’ was initially proposed as a USE reference to the existing ATED term ‘ability’, but closer analysis clarified that students’ capabilities are not seen as fixed abilities but as competencies that can be acquired, which led to the more accurate ‘USE skills’.
Improveing currency

Information technology is an area where both concepts and terminology change quickly, and the Resource Library covers a considerable number of IT-related projects. Related OLT keywords included ‘avatars’, ‘digital immigrants’, ‘digital natives’, ‘learning analytics’, ‘educational data mining’, ‘interactive multimedia’, ‘net generation’, ‘synchronous communication’, ‘Second Life’, ‘emerging technologies’, ‘Web 2.0’ and ‘online assessment’. Quite a number of the terms added to ATED were of this type and improved the currency of the thesaurus as a result. For example, ‘learning analytics’ was deemed to be distinct enough from ‘data analysis’ to be added as a term in its own right.

However, a balance must be struck between currency of the vocabulary and ensuring concepts are sufficiently mature and significant to warrant inclusion. As ATED already carried the terms ‘digital literacy’ and ‘generation gap’, it was decided that both ‘digital immigrants’ and ‘digital natives’ were adequately catered for conceptually; they were added as USE references instead. ‘Second Life’ had been used 48 times in AEI, but it was judged as still primarily the name of a piece of software and therefore out of scope.

Adding terms beginning with ‘online’ inevitably raises the question of whether all educational methods will assume an online component in the near future, making the specification of the ‘online’ aspect redundant. Nevertheless, ‘online assessment’ had been used 44 times as an identifier in AEI; ATED had only ‘online tests’ and ‘student assessment’, and so ‘online assessment’ was duly added.

At first, it was thought that the moment to add ‘Web 2.0’ had passed, but with 587 occurrences in AEI, it really needed to be addressed and was eventually added. On the other hand, although the term ‘personal digital assistants’ had been used 74 times in AEI, as well as by OLT, it was a technology unlikely to return to the fore, and this term was not added.

Synchronising style and syntax

A fundamental constraint of a controlled vocabulary is consistency of style and syntax. The OLT keywords were a mixture of nouns, acronyms and adjectives. When they were brought into ATED they were modified to fit the ATED style of nouns, for instance ‘synchronous’ was changed to ‘synchronous communication’.

Similarly, ‘teaching research nexus’ was changed to ‘teaching research relationship’, in keeping with ATED’s use of the word ‘relationship’.

Broadening consultation

The participation of both academics and librarians on the project team was useful when terms were being scrutinised. While librarians brought their understanding of thesaurus structure to the project, academics’ first-hand knowledge of university practices and context enabled them to contribute definitions for terms like ‘design based research’ (a research methodology used to test the efficacy of educational design). ‘Threshold concepts’ had been proposed as a USE reference to ‘fundamental concepts’, but academics advised that threshold concepts were concepts a student must understand before progressing further and that these could be quite advanced, unlike fundamental concepts. ‘Threshold concepts’ was therefore added to ATED as a new term in its own right.

Broadening sources of authority

No vocabulary exists in isolation, and it is necessary to consult other sources when confirming new concepts and terminology. Aside from the ERIC Thesaurus (the US equivalent of ATED), other sources were used to help in the assessment of terms. The Victorian Education Department confirmed the equivalence of ‘assessment for learning’ and ‘formative assessment’. Exercise and Sports Science Australia distinguished ‘exercise physiology’ from ‘exercise science’. The term ‘chiropractic’ as a noun had to be confirmed by Chiropractors’ Association of Australia. The US National Library of Medicine Subject Headings (MeSH) confirmed ‘histology’ as a subcategory of ‘anatomy’. The Australian Qualifications Framework provided a definition for ‘associate degrees’.

Conclusion

This article has described how a thesaurus was selected and prepared for its use in the re-indexing of the OLT Resource Library. The value of this process was three-fold. The OLT benefited from the maturity of ATED’s structure and terminology, and saved the cost of creating a bespoke vocabulary. At the same time, ATED was strengthened for its use in other indexing, through the testing of the thesaurus against a pool of literature beyond its traditional sources of warrant. The effectiveness of the process was confirmed in the subsequent re-indexing, after which testing showed improved recall and precision rates for the searching of the database. To capitalise on the investment in this project, it will be important for the OLT and ATED’s developers to maintain their close working relationship, so that the thesaurus continues to meet the vocabulary needs of the Resource
Library.

References


Acknowledgements

Support for this project has been provided by the Australian Government Office for Learning and Teaching. The views in this project do not necessarily reflect the views of the Australian Government Office for Learning and Teaching.