2003

Factors influencing the discovery and reusability of digital resources for teaching and learning

Graeme Brownfield

Ron Oliver

*Edith Cowan University*
FACTORS INFLUENCING THE DISCOVERY AND REUSABILITY OF DIGITAL RESOURCES FOR TEACHING AND LEARNING

Graeme Brownfield,
Office of Training and Tertiary Education, Victoria,
brownfield.graeme.w@edumail.vic.gov.au

Ron Oliver,
Edith Cowan University, Western Australia,
r.oliver@ecu.edu.au

Abstract
This paper describes findings and outcomes from the Toolbox Digital Repository project (http://toolbox.flexiblelearning.net.au/index.htm) undertaken in 2002 that was concerned with the design and development of a digital library to support the discovery, access and reuse of the vast number of digital resources developed for online learning settings as part of the National Flexible Toolbox Project. The importance for processes and strategies supporting resource discovery in development stages has emerged from the work as an area needing attention. Digital resources use metadata to facilitate their discovery and recovery. Our project has identified the provision of metadata as an important component of resource development that at this stage appears universally weak and in need of some guidance and support. This paper describes the areas where this process was found to be weak and has suggested some strategies that can be applied at the local level to overcome the problems that were identified.

Keywords
Learning objects, online learning, digital resources, digital libraries, reusability, digital repository

Introduction
Much has been written recently on the topic of reusability as both a design and development strategy for online learning materials and as a general approach to the use of digital resources for teaching and learning (e.g., Downes, 2000). The topic necessarily impacts on the actions of the vast majority of people associated with the teaching and learning process and includes such stakeholders as:

• administrative and financial bodies that look to benefit from the potential cost savings associated with reusing and sharing learning resources;
• policy-makers who are interested in the legal and ethical implications of copyright and intellectual property among the shared objects;
• instructional designers who need to consider design strategies that facilitate and support sharing and reuse; and
• developers who need to consider appropriate development strategies to ensure interoperability and a capability for use of resources beyond the context for which they are designed (e.g., Downes, 2000; Shepherd, 2000).
Added to this list must also be the technical experts who have for some time been looking to describe technical standards and systems for digital resources that:

• facilitate sharing and reuse;
• provide a means for supporting the discovery and access of digital resources that can be shared; and
• guide the development and design of management systems, repositories and libraries which can act as sources and stores of resources for sharing purposes (eg. Wiley, 2000).

This paper describes some findings and outcomes from the Toolbox Digital Repository project ([http://toolbox.flexiblelearning.net.au/index.htm](http://toolbox.flexiblelearning.net.au/index.htm)) undertaken in 2002. This project was primarily concerned with the design and development of a digital library to support the discovery, access and reuse of the vast number of digital resources developed for online learning settings as part of the National Flexible Toolbox Project. The paper seeks to describe lessons learned and knowledge gained from an activity that involved the accumulation of resources in a central repository, the design of system to enable the discovery and access of the resources and a study of the efficacy and utility of the process as a strategy for supporting the reuse of digital resources.

Enhancing the reusability of digital resources
One of the principal factors influencing the reusability of digital resources relates to their capacity to be discovered by intending users. There has been a large amount of research undertaken to develop appropriate forms to develop systems by which digital resources might be described to facilitate their discovery. In the online setting, this work has involved the development of metadata standards, in-text descriptors of important fields, which can be viewed by search engines but which are not evident to the actual users of the resources. A number of metadata schemes have emerged in recent years to provide frameworks to assist in the development of meaningful strategies to support the discovery of digital resources. Perhaps the most well-known metadata standard is the Dublin Core standard that emerged in mid 1990s (DCMI, 1999). While the Dublin Core was among the first set of standards to systematically attempt to provide a system for describing digital resources, many more exist. In the Australian sector, Education Network Australia (EdNA) developed a version of Dublin Core to provide a local context for descriptions (EdNA, 2003) and a version that was consistent with the Australian Government Locator Service (AGLS).

More recently the IEEE learning object metadata (IEEE, 2002) has emerged as a more comprehensive description scheme aimed primarily at supporting descriptions for learning objects. Whereas the Dublin Core metadata initiative proposes 15 descriptors, the IEEE LOM contains 65 descriptors. The effectiveness of such a system is questionable especially when compliance requires developers to use in the least case only one descriptor. To overcome problems with insufficient descriptions, organisations are now mandating a structured use of metadata and requiring application profiles (Campbell, 2003), recommendations that focus on particular elements, descriptions and vocabulary to “facilitate accurate metadata creation and to optimise interoperability between projects” (p.41).

Digital Repositories
The planning and design of the repository for the digital resources in this project involved a number of decisions that had to be taken as a consequence of the data that was to be worked with. Like other settings, the context played a part in the decisions that were made but at the same time, the project was cognisant of the research and development underway that provided guidelines and specifications for work of this nature. There are now a number of groups looking to define standards and to prescribe specifications for digital repositories. The IMS Global Learning Consortium is one of a number of groups promoting open
specifications for supporting e-learning activities. Their specifications currently extend to such activities as locating and packaging educational resources (learning objects) within a learning design framework, tracking and monitoring learner progress and the exchange of the resulting information.

Specifications for digital repositories are a recent addition to the standards activities. A digital repository is a collection of digital resources that can be accessed through a network requiring no prior knowledge of the collection’s structure. Repositories usually hold many forms of digital resource including their metadata descriptors, although the metadata need not necessarily be stored with the various assets. The specifications for digital repositories that are currently being developed by IMS include object querying and locating functions. Recommended standards include the W3C XQuery (2000), W3C SOAP (2001) the simple object access protocol, and ZOOM (2000), the Z39.50 object oriented model. The digital repositories project was aware of these standards and pursued a developmental strategy that recognised the importance of these features while maintaining the flexibility required to deal with the data sets that were to be used, the existing resources and elements from the Flexible Learning Toolbox Project.

Where this digital repository project differs from other contemporary work, lies in the nature of the resources we were looking to store. The resources from the Flexible Learning Toolbox project do not provide a close match in many instances to people’s perceptions of learning objects. In their current form, the Toolbox resources tend to be identifiable at the page level rather than at higher levels of aggregation such as module or topic, the better level for defining and reusing learning objects. Nevertheless we felt that given the scope of the Toolbox resource set, and the possibility of developing technology solutions to overcome some of the aggregation problems, the development of the repository with these resources would strongly support their capacity for reuse.

Flexible Learning Toolbox Project

In Australia, responsibility for national support of the education and training of the Australian labour force is vested in the Australian National Training Authority, a body whose brief is to ensure “that the skills of the Australian are sufficient to support internationally competitive commerce and industry and to provide individuals with opportunity to maximise their potential” (ANTA, 1998a). ANTA’s agenda has been “to deliver training more responsively and efficiently and to a wider catchment area” (ANTA, 1998b). The Authority has long recognised that technological advances have been the catalyst for new forms of educational and training systems and that the demand for, and supply of vocational education and training is increasing. Consequently it has enacted and sponsored strategies to encourage the vocational education and training (VET) providers to support more flexible, technological-based, delivery for domestic and international markets. One of these projects that has grown from an original concept into a well-developed project supporting flexible delivery and learning is the Flexible Learning Toolbox Project.

The concept of a Toolbox as a building block for flexible delivery and open learning is derived from the notion of a set of resources and tools that can be implemented by teachers and trainers to create and implement on-line and computer-based learning environments. The concept is a powerful strategy as a means to create sustainable and scalable materials for technology-based teaching and learning (eg. Hanley, Schneebeck & Zweier, 1998). Since 2000, Toolbox development has been a component of the Australian Flexible Learning Framework (AFL, 2000). In addition to the goal of producing world class online content, there are goals concerning professional development, policy, research, and infrastructure (see http://flexiblelearning.net.au) and there is considerable collaborative effort to integrate these. Policy for the Framework is determined by FLAG (Flexible Learning Advisory
Group). This consists of a representative of each of the 8 states and territories, plus representatives from ANTA, the national Department of Education, Science and Training DEST, and from the ACE (Adult and Community Education) sector. In a decentralised system, a public tendering system establishes which areas are to be covered by a Toolbox and which organisation, or consortium of organisations, will develop it. Although some organisations have developed multiple Toolboxes, the development process is widely spread across the country and includes a large number of development agencies.

The Flexible Learning Toolbox Project showed considerable foresight by making provision of metadata and a metadata standard important elements of the initial project. The Toolbox project required all developers to employ a pre-planned set of fifteen descriptors from the EdNA metadata set. This mandatory requirement appeared to provide a solid basis for any subsequent project supporting the discovery and reuse of the various resources.

While contemporary applications of metadata now involve controlled vocabulary and thesauri, earlier projects, even with discrete description requirements, were less rigid in the application phase and this was a situation that had considerable prospect to impact on the digital repository planned for this project. Because it was expected that each training provider would disaggregate the Toolbox and customise it differently, the metadata was embedded in each HTML page to ensure its continued presence and this affected the level of granularity available for our project. We were also faced with a large number of resources with consistent metadata descriptors but whose content was variable and potentially only partly reliable.

**The Toolbox Digital Repository Project**

The development of a central store of the resources from the Flexible Learning Toolbox Project was an activity that sought to value-add to a very powerful set of learning resources.

The Toolbox Digital Repository Project commenced in February 2002 and sought to explore strategies by which the various files and resources in the discrete Toolboxes might be stored and accessed by teachers looking to discover items that could be used in their own teaching settings. After initial explorations and inquiries, it was decided that an appropriate strategy would be to develop a system whereby the resources could be stored in one location and to compile a database from the metadata contained in each. While the use of metadata alone is not sufficient for resource reusability, metadata tags do allow for the location of resources and since the resources within the Flexible Learning Toolbox Project contained a considerable degree of metadata information, this provided a strong support for any system looking to support resource discovery and access. The software architecture that was chosen to implement the digital repository is shown in Figure 1.
The repository consists of a number of interrelated components that deal with the processing of metadata. The repository contains a copy of each Toolbox CD. Using a ‘robot’ application, every resource within the repository hard drive is checked for metadata. Resources containing metadata are indexed within a database. Once this database is populated, it is ready for searching. A search query entered by a user is translated into an SQL statement that returns a record set from the indexed database. This record set is manipulated and displayed within the web browser for the user to peruse. The architecture described here takes full advantage of the Dublin Core metadata standard used to describe educational resources within Toolboxes.

In its current state of development the Digital Repository contains over 130,000 discrete elements from the 2000 and 2001 Toolboxes. A robot has been developed that has enabled the metadata from these elements to be stored in a database which can be searched using an accompanying search engine. When keywords are entered into the search engine, a page (or series of pages) is produced with links to the various resources that match the keyword search. These pages can then be accessed to review the relevant resources so that decisions about their suitability can be made (Figure 2).
The Digital Repository then provides a means for users to select the items which they would like to reuse. Selected items are added to a collection with the user able to add (and remove) items until all required items have been sourced. When the user is happy with the items and wishes to proceed, on command from the user the system creates a zip file comprising the pages selected and the resources (e.g., images, graphics) displayed on each page, and allows the user to download this file to the user to facilitate reuse of the items in the user’s setting.

The development of this initial stage of the Toolbox Digital Repository was followed by a series of tests and trials including user tests and evaluations of the integrity and accuracy of the resource sets. Whilst the development of the architecture and systems provided a number of technical challenges, these were successfully dealt with in the construction of the site and the resulting system works very much as intended in the design specification. The outcomes from the initial trials and user tests, as well as outcomes from the implementation of the data into the repository have provided a number of findings and insights which may help to inform other projects with similar aims and goals.

Lessons learned

The following sections describe outcomes that have emerged from the initial trials. First, we describe outcomes associated with the capacity of the repository to provide users with a means to adequately discover the contained resources. The bulk of these findings were drawn from a number of limitations discovered in the quality and scope of the metadata that were employed. Then the discussion explores the potential of the site to facilitate reuse of the resources through the processes used to return the resources to the users.

a. scope of metadata

A requirement of the Toolbox project was that developers added metadata to the various resources. Quite stringent guidelines were provided to indicate which metadata elements were to be added. The project used a modified form of the Dublin Core descriptors and developers were required to include 15 metadata fields in each Web page. In instances where multiple pages were contained within frames, developers were allowed to insert the metadata into the frameset rather than having to tag each page separately, a decision that suggested a consideration of granularity was being considered at a modular level.

As developers submitted their completed products for dissemination and application, quality assurance processes were employed to check the various requirements. To check for metadata compliance pages were reviewed in a random fashion across each Toolbox. This process revealed a number of developers were not dealing with the metadata requirements as required and about 10% of the products were returned to the developers for further work on this basis. However, the checking process was found to be extremely onerous and even then did not adequately ensure metadata compliance.

b. metadata quality

Another aspect of the metadata process which was found to be lacking as the Digital Repository project continued was the quality of the metadata descriptors that were being provided for the various resources. It became apparent that in the Toolbox project, different teams employed different processes for the metadata and some processes clearly provided a better outcome than others. For example, in a number of instances the metadata was the responsibility of the programmers and Web page builders. The logic here was these were the people cutting the code and since metadata was code, they needed to be the responsible people. Unfortunately most programmers did not have the capability to
accurately determine the appropriate metadata descriptors for a number of the key fields and often used quite limited descriptors and keywords.

The quality of the metadata was found to vary considerably across the various Toolboxes and even within some products. The most accurate metadata was provided for those fields where the information was most readily available. For example, where metadata was employed, it was most accurate in those fields where the descriptions tended to be fixed across a range of resources. For example, date of development, rights management, developer, etc. The most important metadata for discovery tends to be in the subject and descriptions fields where more open forms of keywords can be employed. It was in these fields where the metadata proved to be least accurate.

c. unique metadata No two resources in any Toolbox were identical so it was expected that no two metadata description sets would be identical. But as the database for the Digital Repository was filled using the descriptors from the various resources, it became apparent that many resources did in fact contain identical metadata descriptors. This may not have been an intentional act on the part of the developers but occurred through as number of likely reasons:
• A lack of skill and capability among those describing the products to adequately describe the content on the various pages;
• Those providing the metadata descriptors were using a limited set of keywords to describe content causing many terms to be reused;
• The lack of any organised thesaurus limited the detail of analysis applied to various pages;
• Limited sets of descriptors being applied to pages, The descriptions were based on too few keywords providing quite broad descriptions rather than fine-grained descriptions as are needed to enable distinctions

d. accessing related resources In the Toolbox project, sets of resources were created comprising many pages. In the discovery process, it was found that teachers would often discover a page from a resource based on their searching but that associated pages would not be found. In particular pages within a topic or module could be found but the pages in the original resource through which they were originally organised, the indices, for example, were not easily sourced limiting the reuse in some respects.

This problem was caused by the flexibility adopted in the Toolbox project for the grain size of discoverable resources. In the quest to make pages more discoverable, the system had in fact removed some of the important contextual information contained within sets of pages and not evident when individual resources were discovered. One likely solution to this problem appears to be in the more deliberate use of the Relation metatag. The DC.Relation descriptor provides a means for developers to indicate relationships between various pages. This strategy provides a means to have flexible sizes of the digital resources and enables related pages to be discovered once a resource has been found to be useful. Once again, for this strategy to be employed effectively, it would be necessary to adopt a consistent format for nominating the relationships and describing related pages.

e. providing metadata for all resources Another granularity issue that arose in explorations of the resources within the Digital Repository and supporting their discovery was the potential advantage and value of having metadata for all discrete media elements including images, movies, graphics, sound files. While many would argue that such a strategy would create far more accessible resources and provide an efficient means for supporting wider levels of reuse, our experience suggested that this was not necessarily a sensible strategy.
There were a number of factors that suggested the notion of including metadata descriptors within the media elements was not likely to support discovery in any meaningful way:

- Many of the medias elements served as supports for layout, interface and navigation and the need to adequately describe these elements would be a large task and serve little useful purpose;
- All media elements in the Toolboxes sat necessarily within a larger resource, for example a Web page. Providing the Web page was adequately described, a teacher looking for resources and using appropriate keywords would be able to find the media elements through the search of the larger resource, from which it could then be extracted if it could be used elsewhere;
- Many media elements can only be useful in their applied context. Sound files for example of descriptions and feedback are nearly always associated with some particular activity and supported by textual and graphical elements. Discovering them outside this context would render them almost useless since very few can actually be used independently of the context in which they were originally planned.

f. weighting and ordering of search results  As is common with most search engines and large databases of information and resources, many searchers will return a large number of hits leaving users with the problem of deciding which resources might best suit their needs. It became apparent in this project that our repository needed to have some automated strategy for sorting and selecting resources from keyword searches that enabled a more purposeful ordering for the user rather than a random order or sequence dependent on position of records in the database.

Many possible solutions appear to exist which could provide some meaningful ways to assist an automated discovery process to order resources and these will be considered in further revisions of our processes and refinements of the search engine. Without some formalised process that enabled the searching algorithm to calculate relative importance, any searching process being used across such a large number of resources would always appear to have some inherent weaknesses limiting aspects of its usability and utility.

g. catering for resource context  The results from our searching trials with the digital repository demonstrated a number of successes in terms of being able to identify and access many discrete resources. But it was the discreteness that appeared to suggest a need for some further thinking in our design strategies. Many of the individual resources that were appearing in the results window were resources that came from modules and larger organisational learning sequences. Discovering them discretely was a very useful outcome but it was apparent that there needed to be some way to link them to their large context. A number of possible strategies are now being actively pursued within the project to facilitate this process.

One solution that appears to offer considerable potential is that of providing the resources in a framed window enabling users to explore the various resources associated with the discovered resource so that related resources that are found to be useful, may also be accessed. The repository in our project provides a number of choices to the user and this format seemed to be in keeping with the design philosophy of providing user choice while at the same time using the technology wherever possible to facilitate these actions. This frames approach should make the discovery process easier for users than other options.

h. interoperability with other repositories  As work progressed with this repository project, we became aware of a variety of other similar projects being undertaken in the local context. There are many organisations looking for discovery and access solutions for their digital learning resources and we became cognisant of the need for these activities to have some
form of consistency. It would seem very important for any work to have the flexibility to enable connections with outside projects as well as being internally consistent.

In our experience we observed several projects where local features, such as file naming protocols and storage strategies were inconsistent with our own. We initiated some dialogue to discuss considerations of consistency and standards. The emerging standards for digital repositories including the W3C querying and accessing protocols are very useful starting points. The important point that emerged from these communications was the fact that within the local higher education educational setting we will probably soon have a large number of disconnected digital repositories and there will be some projects seeking to find ways to provide consistent access between them.

**Strategies supporting reusability of digital resources**

The outcomes from this work lead us to suggest a number of strategies and processes could be employed relatively efficiently in the development processes of digital resources for learning settings that could contribute significantly to their capabilities for reuse. In particular we feel that the area where the most pressing work is needed is in support of the metadata processes. In any projects where reusability will be an important factor, the value of accurate and reliable metadata cannot be overlooked. We propose the following as some possible strategies for achieving this aim.

**a. thesaurus** There is a need for developers and teams to have access to a thesaurus so that any future descriptors that are chosen for metatags will be consistent and accurate. There is a need to provide a thesaurus for use with training materials that includes the use of industry (work) based terms as distinct from the descriptors and terms which result from the research, academic and project reports more commonly at present in collections. To assist the development of such a thesaurus, we have provided a list of all the DC.Subject keywords currently used in about twenty of the Toolbox products to illustrate the needs of the vocational education and training sector.

**b. identifying metatagging experts** The provision and addition of accurate and reliable metadata to digital resources is a process that requires a number of particular skills and capabilities. Currently there seem to be few people with these skills and many novices who are needed to develop this competence in a very short period of time. Our project has identified a number of people with these skills who are now in strong positions to lead and guide others. Our project suggests that there would be many advantages and opportunities to be derived from the identification of metatagging experts and initiating their involvement across a series of projects.

**c. sharing metatagging expertise** The metatagging process has now become an important component in large scale online resource development and this project has demonstrated the important need for this expertise in any development team. The use of an identified metatagging expert across projects provides a number of opportunities and advantages. This person could be a virtual member of the teams due to the electronic nature of the work with services shared between teams. Such an approach would help to ensure the integrity of the metadata within any single large scale project but would have the added advantage of ensuring consistency in metadata between connected projects. Currently there appear to be few people with the necessary capabilities and this form of sharing has the prospect not only to improve the metatagging process in current projects but also of providing an effective way to train and develop others.
d. Adding metadata at the repository stage  Metadata is a commodity that can have many consistent features within projects and it is a commodity that can be added at any stage in the development process. In the Digital Repository project, we found it necessary to add several extra fields at the storage stage and it became apparent to us that there was other common metadata (e.g., copyright and publisher) that could be added at the final stage rather than at the resource development stage. The notion of providing some form of automated metatagging process could provide a number of flexibilities supporting reuse of the resources while assisting to ensure the accuracy and integrity of the descriptors. Already there are a number of tools that can be used to retrospectively add metadata to digital resources (e.g., EdNA, 2003) and the use of such tools within projects should be explored further to determine how best they might be employed.

e. documenting management strategies  At the moment the process of providing metadata to digital resources is very much seen as a peripheral activity to the development process. It would appear that this activity needs to be given mainstream status and measures adopted to see it become more of an integral than additional activity. There seems to be a need for some form of documentation of strategies and approaches that can be undertaken in projects to ensure that the metatagging is carried out in an efficient and organised fashion. Our project has suggested that value could be gained from having successful projects document their processes as a guide for others and that this process be explored and developed into a formalised and well-articulated form.

The existence of the Digital Repository is already having a positive effect in several of the areas previously discussed. During the induction workshops for the 2002 and 2003 Toolbox projects, developers were quick to recognise its value for users of their products. Metadata provision now appears to be recognised as a mainstream part of each project rather than as a pedantic contractual obligation. Some teams are building the collection of keywords and descriptions into the storyboarding stage under the guidance of instructional designers. Others are employing a specialist to categorise the pages and add metadata as components are completed.

Through the Digital Repository project, a small software tool was created based on the design of the robot. This tool enables the resources within a Toolbox to be checked for metadata through an automated process. All new developers have been given the tool and asked to check their products. This process places the onus back on the developer to demonstrate compliance and provides some further assurance that metadata will be included in the pages and resources that are now developed.

**Summary and Conclusions**

The outcomes from this project have been very positive and provide us with the expectation that the Digital Repository we have created will serve to support the reusability of the resources in the National Flexible Toolbox project. The intention now is to load the site with the resources from all the completed projects which will add several hundred thousand more resources to the database. Our research will turn more now to exploring how teachers and trainers are able to use these resources and what accompanying features are needed to support and encourage teachers to make use of the resources in their teaching and learning.

In terms of what we have learned and the forms of advice and action that will likely filter to the current Toolbox Projects, the need for processes and strategies supporting the discovery of resources has emerged from our work as the area needing most attention. Digital resources use metadata to facilitate their discovery and recovery. Our project has identified the provision of metadata as an important component of resource development that at this stage appears universally as a weak process and in need of some guidance and support.
This paper has described the areas where this process was found to be weak and has suggested some strategies that can be applied at the local level to overcome the problems that were identified. These strategies would likely be useful in any such project in any setting. In this way the findings from this project might serve to inform and guide other projects exploring reusability, including those where there is a stronger technical imperative.

The findings from the Digital Repository project will be used to guide subsequent implementations of the Flexible Learning Toolbox project. As this information is fed into the Toolbox project and used to inform subsequent activities, we intend to continue our research and to make some detailed explorations of usability questions in terms of how teachers use the repository and what learning advantages such resource sets offer to teachers in the mainstream forms of teaching.

References


Acknowledgements
The authors would like to acknowledge the contribution of the Edith Cowan University Research Office and the Office of Training and Tertiary Education who provided the funding to this project.

Copyright [] 2003 Graeme Brownfield & Ron Oliver

The author(s) assign to ASCILITE and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author(s) also grant a non-exclusive licence to ASCILITE to publish this document in full on the World Wide Web (prime sites and mirrors), publication to CD-ROM and in printed form within the ASCILITE 2003 conference proceedings. Any other usage is prohibited without the express permission of the author(s).