

Considerations for a Digital Repository

R. M. Wirski¹, R. Oliver¹, P. Hingston¹, A. Omari¹, and G. Brownfield²

¹ *Edith Cowan University*
Perth, Australia

Email: (r.wirski, r.oliver, p.hingston, a.omari)@ecu.edu.au

² *Flexible Learning and Infrastructure Unit Learning Technologies*
PO Box 266D Melbourne Vic 3001

Email: brownfield.graeme.w@edumail.vic.gov.au

Abstract

This paper describes issues surrounding the role of metadata within educational resources. The use of metadata is used within a range of electronic information systems to provide information about the content of electronic resources. The paper discusses the use of metadata within learning objects and discusses their potential for increasing the reusability of educational resources. The paper also discusses other systems currently being explored to support reusability of learning resources including SCORM. We provide a brief outline for a digital repository to support the discovery and retrieval of the resources developed through the National Flexible Learning Toolbox Project.

1. Introduction

The concept of learning objects is fundamental to such issues as the reusability and customisation of resources and digital rights management. There is a high degree of initiative and activity in the field of learning objects within Australia and the purpose of this paper is to describe aspects of one such project, the Digital Repository project, a joint effort between Edith Cowan University and the Australian National Training Authority. This project is centred on exploring strategies to promote the reuse of resources from the National Flexible Learning Toolbox Project. The key component of this project is an exploration of ways to collectively store, discover and retrieve resources from the Toolbox Project to support their potential customisation and reuse.

The National Flexible Toolbox Project has overseen the design and development of a large number of online resources to support the flexible delivery of a number of qualifications within a range of Training Packages. The Project has sought to generate these resources in ways which support customisation and reuse and has employed a number of strategies to foster these aims. Among these strategies has been the deliberate use of metadata within the development process. This paper discusses: current metadata standards; the relevance of learning objects as a strategy supporting discovery and reuse; and the progress of the joint

ECU-ANTA project exploring ways to store, discover and retrieve digital resources from the Toolbox Project.

2. Learning Objects and Education

With the proliferation of electronic content supporting Web-based learning, many people often question why there is so much constant duplication of resources. The nature (and essence) of the WWW is such that it provides very strong support for interoperability and the sharing of resources. For example, a teacher preparing a set of learning materials for a course needs only to use a hyperlink to include materials from another teacher among the resources. The reason why developers tend to build new resources before they look for sharable resources tends to revolve around the need to have some form of control and influence over one's materials. Widespread resource sharing without consistent systems makes this a difficult goal to achieve.

Enter learning objects. The learning object is a relatively young concept in educational technology. A learning object, as defined by IEEE's Learning Technology Standards Committee is "any entity, digital or non-digital, which can be used, re-used or referenced during technology supported learning." The main characteristics of a learning object are reusability and metadata tagging.

Reusability is the biggest benefit (and motivation) in using learning objects. Learning objects are characterised by the belief that we can create independent components of educational content that provide an educational experience. The learning object viewpoint states that these components are self-contained, though they may contain references to other components; and they may be combined or sequenced to form complete courses. The advantages from design strategies that incorporate the use of learning objects stems from the ability of such processes to provide libraries of discrete digital resources that can be used in flexible ways to support technology-facilitated learning in its many forms

Longmire (2000) suggests a number of distinct advantages can be achieved from design and development strategies which revolve around learning objects. These include:

- A capacity for materials to be made and used more flexibly;
- Support for updating and maintaining learning resources;
- Support for customisation of resources to meet the needs of different users;
- A capacity for interoperable use across a variety of systems; and
- Increased value of content as a consequence of its ability to be used in more than the original setting.

When meeting learning objects for the first time, people often wonder what they might look like. A number of writers have suggested different strategies for describing objects. Koppi & Hodgson (2001) describe 5 classes of learning object ranging from a raw asset, an object with inherent educational context, for example a picture, through learning assets, tasks to content free templates supporting learning such as role playing engines. The vast differences between the various items that constitute learning objects adds to the complexity of the processes needed to define and describe them.

3. Metadata Standards

An important requirement for the identification of learning objects is the application of metadata. For learning objects to be used intelligently, they must be described using an appropriate scheme, in much the same way as books and resources are indexed in libraries. With electronic sources, the most common form for indexing elements is through the use of metadata systems. Metadata standards such as the Dublin Core and EdNA can be used for this purpose.

There are several leading standards that focus on descriptive metadata that relate to educational uses. These are:

- Dublin Core MetaData Initiative.
- Education Network Australia (EdNA).
- Australian Government Locator Service (AGLS).
- Canadian Core Learning Metadata protocol (CAMCORE).

The principal purpose of these metadata standards is to facilitate the aggregation of metadata about educational resources. Typically this is achieved through the use of a set of guiding principles and elements. Each element has an associated description, and these descriptions provide the distinction between various educational resources.

Metadata within these standards is often stored as name-value pairs within meta tags, which are placed within the “HEAD” elements of a HTML document. However, it can also be located in an external document (e.g. an XML file) or loaded into a database, enabling it to be indexed and manipulated from within a proprietary application.

An example of metadata for Edith Cowan University’s SCIS home page is:

```
<link rel="schema.DC" href="http://purl.org/dc">
<meta name="DC.Title" content="SCIS Support">
<meta name="DC.Publisher" content="Edith Cowan University">
<meta name="DC.Date" scheme="W3CDTF" content="2001-05-02">
<meta name="DC.Type" scheme="DCMIType" content="Text">
<meta name="DC.Format" content="text/html">
<meta name="DC.Format" content="583 bytes">
<meta name="DC.Identifier" content="http://www.scis.ecu.edu.au/">
```

The first line defines the metadata standard used – in this case it is the Dublin Core standard. The subsequent lines are name-value pairs that provide some description about the resource. Most metadata standards can be stored using an external XML document, or embedded within the HTML document being described. An example of an XML file storing the same metadata in the Dublin Core format is shown here:

```
<?xml version="1.0"?>
<DublinCore>
  <Title>
    SCIS Support
  </Title>
  <Publisher>
    Edith Cowan University
  </Publisher>
  <Date>
    2001-05-02
  </Date>
  <Type>
    Text
  </Type>
  <Format>
```

```
text/html || 583 bytes
</Format>
<Identifier>
  http://www.scis.ecu.edu.au
</Identifier>
</DublinCore>
```

3.1 Relationships Between Standards

The Dublin Core Metadata Initiative is widely recognised as one of the original metadata standards in use today. EdNA, AGLS and CAMCORE have been derived from the Dublin Core standard.

Most Dublin Core derivatives maintain its fifteen elements, and add new elements that are designed to meet the specific needs of a particular community. For example, EdNA created new standards for the Australian education and training community, while AGLS focuses on Australia's government organisations.

4. Identifying Learning Objects

The capacity for discovery and reusability represent significant benefits for learning objects. To fully realise these benefits, it is necessary address a number of issues: a standard/language for specifying learning objects; learning object size; and a metadata standard.

Reusability is not an absolute goal – it can be achieved in varying degrees. A learning object may be considered reusable if the authoring system used to create it can combine the learning object more than once, while creating learning sequences. The capacity for reuse of learning objects is increased when the objects are designed and developed in consistent and standard ways. There is considerable effort being undertaken today to define possible standards for learning objects that can be applied universally. The application of universally accepted standards in the design and development of learning objects will enable true reusability to be achieved and will facilitate the design of automated systems that can assist in the development of learning settings based on the use of learning objects from digital repositories and libraries.

Learning object size is another important issue. To some degree, the smaller the size of a learning object (in terms of information), the greater the likelihood that the object will be useful to someone else and reused. But, if learning objects are too small then there is a danger that they will become meaningless and not reusable. For this reason, there are a number of classification schemes that provide more meaningful contexts for describing the granularity of discrete objects.

In the National Flexible Toolbox Project, developers have been encouraged to create their digital resources in ways which enable the various elements within the various qualifications to be disaggregated and separated. This strategy was chosen to provide a means for the smaller elements that make up full online courses to be identified and for the constituent elements to be discoverable. For example, in a module that was designed to teach communication skills, it was intended that the various files in the form of: HTML pages of information and instruction; diagrams in the form of gif images; and perhaps the self-tests in the form of Flash movies, would all reside as discrete files and developed in ways that might enable them to be used again in other settings if it was felt appropriate.

To discover, retrieve and prepare such learning resources for reuse can be quite a complex task. In the National Flexible Toolbox Project, the aim has intentionally been simply to

support the possible reuse of the various resources that have been developed. Any attempt to actually reuse the resources will require considerable work on the part of developers and designers. For those groups seeking to fully explore the potential of learning objects as building blocks for online learning materials, their aims are much higher. A number of groups are developing design standards and protocols that will enable the technology to seamlessly integrate objects built for one course into another so that teachers with few technical skills can create diverse learning settings by simply selecting the resources they need and having technology connect them into a learning module.

These are the aims and intentions of such groups as the IMS Project and the SCORM which are seeking to provide very detailed standards and guidelines for learning objects. While the IMS specification was perhaps the first to provide a description of prototype learning objects, SCORM now provides a more specific and, a specification for learning objects that is growing in use and application.

5. SCORM

SCORM is short for the Sharable Content Object Reference Model. It has been developed by the Advanced Distributed Learning (ADL) initiative, which was created by the United States Department of Defence.

The initial motivation was to develop a strategy for using learning and information technologies to modernize education and training and to promote cooperation between government, academia and business to develop e-learning standardization.

5.1 Description

The SCORM applies current technology developments – from groups such as the IMS Global Learning Consortium, the Aviation Industry CBT (Computer-Based Training) Committee (AICC), the Alliance of Remote Instructional Authoring & Distribution Networks for Europe (ARIADNE) and the IEEE Learning Technology Standards Committee (LTSC) – to a specific content model to produce recommendations for consistent implementations by the vendor community.

It is important to note that the scope of SCORM is not all-inclusive. SCORM, although already accepted by industry leaders, will be extended in scope.

SCORM consists of two major components as shown in Figure 1:

- Content Aggregation Model (CAM); and,
- Run Time Environment.

The Content Aggregation Model relates to the creation and storage of educational content while the Run Time Environment describes universal methods for accessing this content.

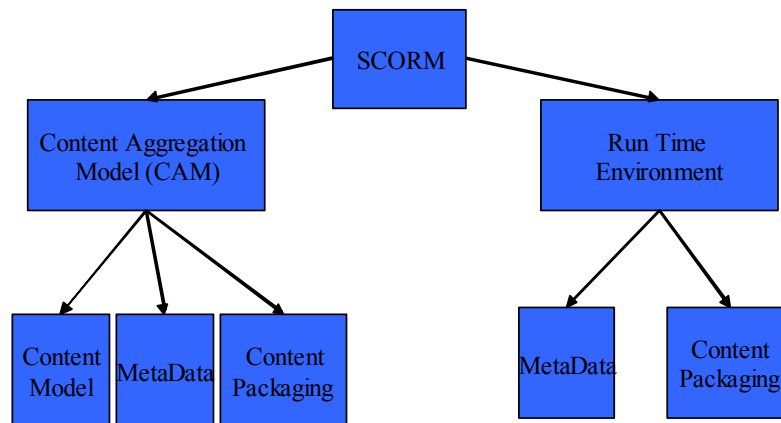


Figure 1 : Overview of SCORM components

Content Aggregation Model

The SCORM Content Aggregation Model provides a means for designers and implementers of courses to combine learning resources for the purpose of delivering a desired learning experience. It is made up of the following:

- Content Model: Nomenclature defining the content components of a learning experience. The content model is the part of SCORM that most closely relates to learning objects through the use of Sharable Content Objects (SCOs).
- Metadata: A mechanism for describing specific instances of the components of the content model.
- Content Packaging: Defines how to represent the intended behavior of a learning experience (Content Structure) and how to package learning resources for movement between different environments (Content Packaging).

Content Model

The SCORM Content Model describes the SCORM components used to build a learning experience from reusable learning resources, i.e. learning objects. The Content Model also defines how these learning objects are aggregated to compose courses. The SCORM Content Model is made up of the following components: Assets, Sharable Content Objects (SCO) and Content Aggregations.

Assets

SCORM learning content in its most basic form is composed of Assets that are electronic representations of media, text, images, sound, web pages, assessment objects or other pieces of data that can be delivered to a Web client. An Asset can be described with Asset metadata to allow for search and discovery within online repositories, thereby enhancing opportunities for reuse.

Sharable Content Object (SCO)

A Sharable Content Object (SCO) represents a collection of one or more assets. A SCO represents the lowest level of granularity of learning resources that can be tracked by a Learning Management System (LMS) using the SCORM Run-Time Environment. SCOs are intended to be subjectively small units, such that potential reuse across multiple learning objectives is feasible. The SCORM does not impose any particular constraints on the exact size of a SCO, although it implies that the content developer will determine the size of the SCO based on how much information is needed to achieve the learning outcome and on the level of reuse that the content developer wishes to obtain.

Content Aggregation

SCORM Content Aggregation is a map (Content Structure) that can be used to aggregate learning resources into a cohesive unit of instruction (e.g. course, chapter, module, etc.).

Run Time Environment

The three components of the SCORM Run-Time Environment are Launch, Application Program Interface (API) and Data Model.

The Launch mechanism defines a common way for a Learning Management System (LMS) to start Web-based learning resources. This mechanism defines the procedures and responsibilities for the establishment of communication between the delivered learning resource and the LMS. The communication protocols are standardized through the use of a common API. The API is the communication mechanism for informing the LMS of the state of the learning resource (e.g., initialized, finished or in an error condition), and is used for getting and setting data (e.g., score, time limits, etc.) between the LMS and the Sharable Content Object (SCO).

A Data Model is a standard set of data elements used to define the information being communicated, such as the status of the learning resource. In its simplest form, the data model defines elements that both the LMS and SCO are expected to “know” about. The LMS must maintain the state of required data elements across sessions, and the learning content must utilize only these predefined data elements if reuse across multiple systems is to occur.

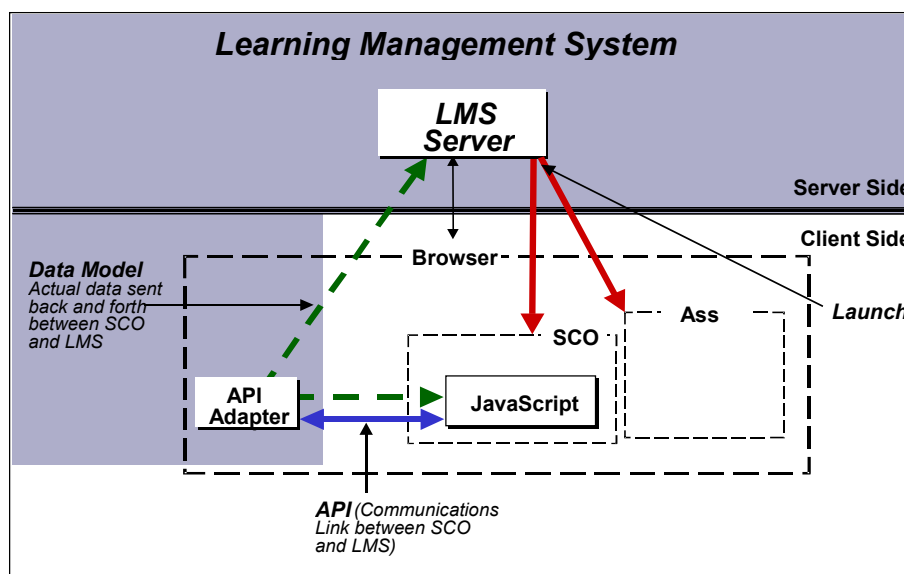


Figure 2 : SCORM Run Time Environment

A full SCORM implementation will allow:

- Course creation by combining reusable learning objects.
- The ability to search for specific courses, learning objects or media through the use of metadata tags at low levels.
- Courses to be used in courseware delivery systems such as Blackboard, WebCT or Click2Learn's Toolbook.

There are a number of drawbacks in adopting the SCORM model for the creation of educational resources:

- The end user must have access to software that can present SCORM compliant courses.
- The implementation of a SCORM compliant system will be a complicated and expensive process.
- The administrative problem of hosting an in-house SCORM compliant LMS system.
- SCORM is not a finished standard.

6. Project Discussion

While the use of metadata alone is not sufficient for resource reusability, metadata tags do allow for the location of resources. Content within the National Flexible Learning Toolbox Project contains metadata information, adapted from the EdNA description of the Dublin Core elements. This metadata has been used as the basis for a digital repository.

It is important to make a distinction between the effectiveness of a search engine and the potential effectiveness provided by metadata usage. When metadata is used in the development of digital resources, it provides a means to identify and retrieve resources based on keyword searches but the effectiveness of the search is based very much on the quality of the metadata and the accuracy of its use. One of the activities in this will be to evaluate the metadata usage within the Toolbox Repository resources.

The software architecture we have chosen to implement the digital repository for the Toolbox Digital Repository Project is shown in *Figure 3*. The repository consists of a number of interrelated components that deal with the processing of metadata.

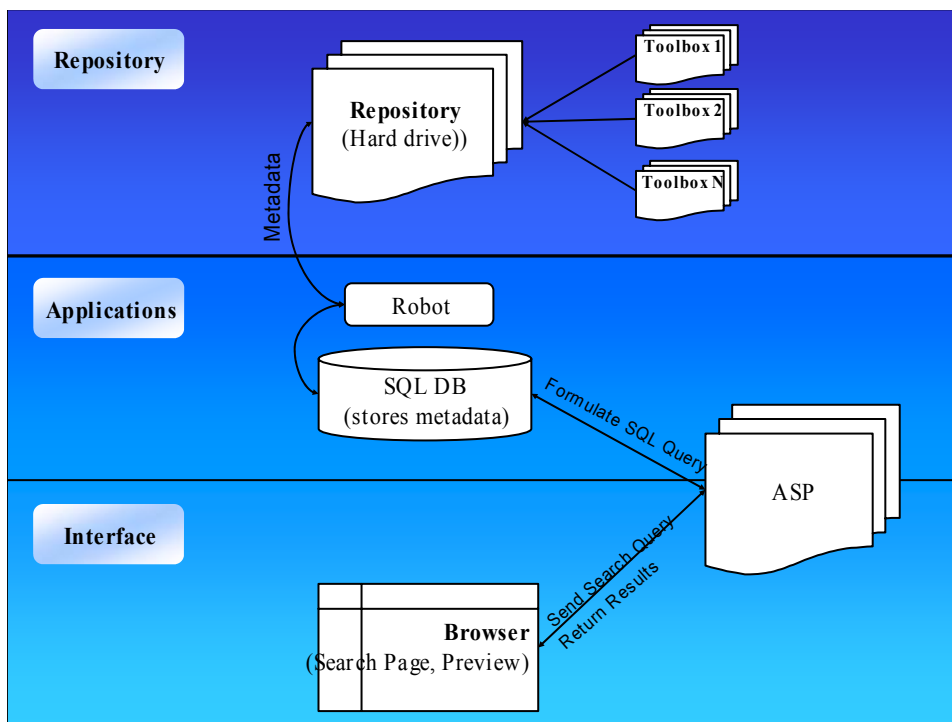


Figure 3 : Toolbox Digital Repository Architecture

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 EdNA metadata standard used to describe educational resources within Toolboxes. Development of this system is continuing, while efficient retrieval of resources has already been demonstrated.

In its current state of development, the Digital Repository contains over 130,000 discrete elements from the Series 2 and Series 3 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 (Figure 4).

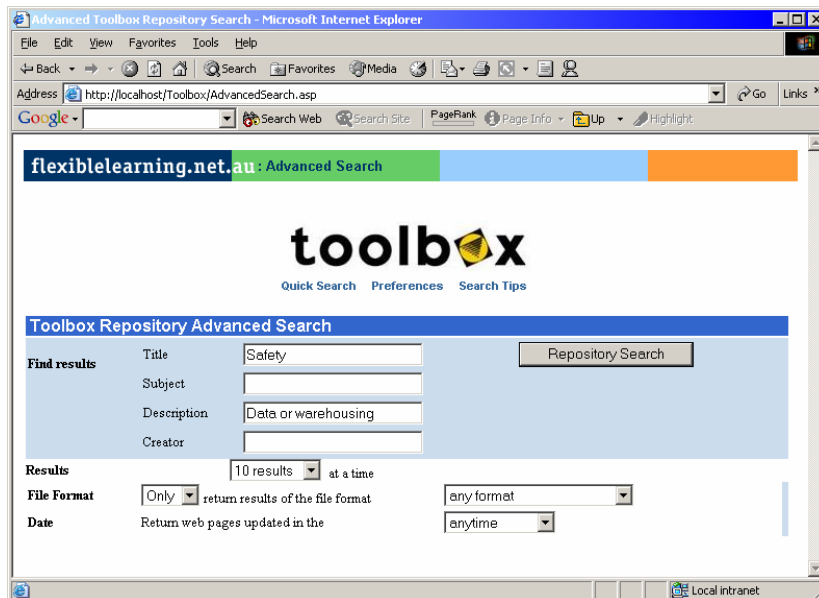


Figure 4: Toolbox Digital Repository Search Engine

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 5).

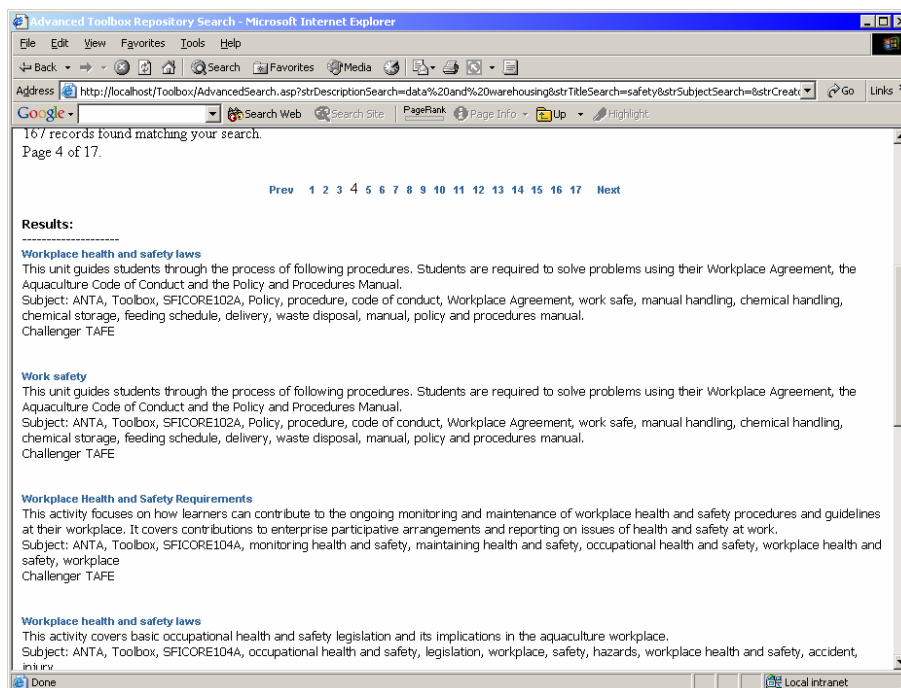


Figure 5: Search Results for the Toolbox Digital Repository

The planned next stage in our Digital Repository project will be to investigate strategies by which resources within the repository can be successfully accessed through keyword searches and their contents viewed by the user. We are seeking to explore the possibility of devising a means by which resources that are selected can be downloaded from the Repository for use in a chosen setting. Challenges that we face in achieving this aim include:

- Developing ways to ensure that resources and all their dependent files and links can be identified and grouped meaningfully;

- Automating the process so that any authorised user can undertake the process through a Web link;
- Planning a strategy to enable rights management to be facilitated; and
- Explorations of effective ways for trainers and RTOs to use the system to ensure full advantage is taken of the various resources and their potential for re-use.

7. Conclusions

The expectation that any resource containing metadata tags will be reusable cannot always be met. For a whole resource, or any subcomponent of that resource to be reusable, it must be produced in a way that facilitates its use by others. This is the motivation for the current work being undertaken to explore the definition of standards for the specification of learning objects. Current standards such as the SCORM model are now providing explicit guidelines that support not only the storage and retrieval of learning objects but also provide support in the area of content packaging. This makes it a very useful framework for creating reusable learning objects.

In the context of this project, the discovery and retrieval of the digital resources within the National Flexible Toolbox Project is based very much on the disaggregated design strategies employed in their development and the use of the EdNA metadata standards across each discrete resource. Our project has involved the development of a robot to extract the metadata information from each resource and to store this information in a database. Searching the database using keyword searches enables the user to discover the various resources and to view them as a means to explore their potential for reuse. The National Flexible Toolbox Project has intentionally sought to use an approach to the design and development of digital resources that will facilitate their reuse in a variety of delivery systems. Future activities in the project are aimed at exploring ways to support the retrieval of the resources to facilitate their reuse in online and technology-facilitated learning settings.

The long-term goal of this work is to help the National Flexible Learning Toolbox project achieve reusability of its educational resources. The project will seek to outline the processes and strategies needed to ensure the reusability of future resources and provide a framework for this.

References

Dodds. P. (editor) (2002), *SCORM Version 1.2*. Reference Document prepared by the Advanced Distributed Learning Team. (www.adlnet.org).

Duncan C. & Campbell L. (2002), *Using Metadata in Packaged e-Learning Content*, presented to UK-MEG, 18 April 2002.

Hillman D. (2002), *Using Dublin Core*, (<http://dublincore.org>)

Koppi, T. and L. Hodgson (2001). *Universitas 21 learning resource catalogue using IMS metadata and a new classification of learning objects*. Ed-Media 2001, Finland, AACE.

Longmire, W. (2000). *ASTD Learning Circuits: A primer on learning objects*. Available at: <http://www.learningcircuits.org/mar2000/primer.html> [Accessed July 2002]

Richvalsky J. & Watkins D. (1999), *Design and Implementation of a Digital Library*, The Fourth ACM Conference on Digital Libraries.

Sokvitne L (2000), *An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval*, State Library of Tasmania.

Wu C (2000), *Distributed Metadata Systems and Retrieval Error Ratio*, Proceedings of 1999 EBTI, ECAI, SEER & PNC Joint Meeting