This is the published version:

Currie, Michael, Moss, Nicholas, Ip, Albert and Morrison, Iain 2000, The EdNA metadata toolsets : a case study, in The Web : communication and information access for a new millennium : proceedings of AusWeb2K, the Sixth Australian World Wide Web Conference, Southern Cross University, Cairns, Qld

Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30046143

Every reasonable effort has been made to ensure that permission has been obtained for items included in Deakin Research Online. If you believe that your rights have been infringed by this repository, please contact drosupport@deakin.edu.au

Copyright : 2000, Southern Cross University
The EdNA Metadata Toolsets: A Case Study

Michael Currie, EdNA Higher Education Project Manager, Dept of Information Systems, University of Melbourne, Melbourne, Australia  m.currie@dis.unimelb.edu.au

Nicholas Moss, Principal Consultant, CRC for Enterprise Distributed Systems Technology (DSTC)  nickm@dстc.com

Albert Ip, EdNA Technical Specialist, Dept of Information Systems, University of Melbourne, Melbourne, Australia  albert@dls.au.com

Prof. Iain Morrison, Dept of Information Systems, University of Melbourne, Melbourne, Australia  i.morrison@dis.unimelb.edu.au

Abstract

In 1999 the EdNA Higher Education project team was funded by DETYA to produce a suite of metadata tools for educational sectors and institutions in Australia. These tools were aimed at simplifying and partially automating the creation and maintenance of metadata for web resources intended to be added to EdNA Online. They also aimed to improve search effectiveness through linking an educational thesaurus to keywords obtained from the resource.

This paper describes each of the projects and the metadata tools that have been produced. These include a Metadata Manager, a Metadata Editor and a linked Educational Thesaurus. The tools provide a range of benefits to resource managers and creators in indexing and enabling access to their resources.

Commencing as three separate projects, the paper describes the process that resulted in a single integrated set of tools being produced. This process provided a learning experience for the project team and for the software developers, DSTC. In attempting to provide best practice in a cutting edge environment, the project became an evolutionary process that aimed to match perceived and surveyed needs to possibilities.

Having successfully produced the tools, the challenge of distribution and promotion to a highly diversified and widely spread clientele had to be met. This involved developing efficient methods of distribution and effective training programs.

The paper discusses some of the lessons that were learnt including the development and testing of cooperative frameworks involving different educational sectors and contract developers, management of resources and vision discernment. Implications and possibilities for future developments are discussed and an analysis of the process and outcomes of the projects is provided.

Introduction

In February 1999, the first author accepted the role of EdNA Higher Education Project Manager based at the University of Melbourne. One of the responsibilities of the position was the management of three projects funded as part of the Framework for Open Learning Program. These involved the development of a set of online tools to facilitate the creation and management of metadata by EdNA stakeholders and the development of an associated thesaurus to improve the search efficiency of the tools. The following twelve months became a learning process that has finally resulted in the production of cutting edge software that has generated interest both nationally and internationally.

This paper aims to give an overview of the functionality and application of these tools and to give an insight into issues relating to the development of resources for a highly distributed and composite audience.

EdNA Online

To understand the reasons behind the development of the tools, it is useful to know something of Education Network Australia (EdNA) and its role. EdNA is the result of a unique collaboration between all sectors of education within the Commonwealth and States of Australia. Initiated in 1995, EdNA provided a forum in which officials from education systems and sectors could collaborate including the sharing of developments and leveraging from each other's experience. Through EdNA, authorities explored the possibility of pooled IT purchasing, of joint Internet activity and the development of common standards. From this emerged the primary aim of optimising the benefits of the Internet for education.

EdNA has a highly developed collaborative structure, which operates in both a hierarchical and distributed manner. While it has a high level governing body, the EdNA Reference Committee, and corporate management, Education.Au, it also has ground level representation through the various sectoral advisory groups.

Central to EdNA's functionality is the EdNA Online website. This features a comprehensive directory of resources relating to Australian education containing some 9000 evaluated sites and over 250 000 linked sites. These resources have been collected from stakeholder collections and from international sources by Directory Officers and through local contributors.

Opportunity

EdNA Online provides users the choice of a Browse or Search functionality in accessing resources. While initially resources were manually organised into a browse hierarchy, the sheer number of resources made an alternative search strategy desirable. After reviewing various metadata schemes and the increasing use of RDF (Resource Description Format), a team from EdNA that included Jack Gilding and Jon Mason set out to develop a metadata specification suitable for Australian educational resources. They adopted the fledgling Dublin Core standard and added elements specific to the needs of the EdNA...
This became the EdNA Metadata Standard [HREF6], Version 1.0, which was launched in August 1998. It incorporated the fifteen DC elements as well as a further nine tags that enabled the resources to meet EdNA quality standards and to be linked to numbered categories in the EdNA directory. Currently a working group is producing Version 2.0 of the Standard that will include more education-related elements in line with developments within the Dublin Core Metadata Initiative.

As the number of useful resources on stakeholder sites continued to grow, it was recognised that a more diversified and automated process for the creation and management of metadata was required. By making it easier to create and maintain metadata at the institutional level, it was felt that not only would it reduce the load on Directory Officers attempting to do this centrally, but it would encourage resource creators and those with an intimate knowledge of the material to include tags which would reflect their greater understanding of the concepts within the resource.

Prof. Iain Morrison, then DVC(IT) at the University of Melbourne, proposed that existing metadata tools might be enhanced to meet the needs of different stakeholder groups. These required that the tools be simple to learn and use, could run on stakeholder sites using a variety of platforms and be capable of being updated by EdNA personnel. It was necessary that the tools not be resource hungry or too cumbersome to download, yet have sufficient computer intelligence to improve efficiency and standardisation of metadata creation through partial automation of the process. It was also important that, in creating metadata, the tools adhered to the quality standards required of evaluated resources on the EdNA directory.

The proposal to develop the metadata tools was also strongly influenced by work on a related project designed to selectively collect evaluated resources from stakeholder sites. The Harvest Project utilises a Harvest Control List (HCL) of metadata-enabled sites to direct a harvest robot to selected resources for adding to the EdNA directory. However it relies on resources already having EdNA approved metadata. It was envisaged that the metadata tools would facilitate the process of adding this metadata and be able to link newly enabled sites to a HCL for harvesting.

Objectives

From discussions with stakeholders, three desired outcomes were identified. These were the creation and management of metadata for individual resources based on the implementation or modification of existing tools, the retrospective creation of metadata for existing unevaluated resources and the desire to improve searchability of the EdNA directory through the controlled use of keywords.

Subsequently three proposals were submitted to the Department of Education, Training and Youth affairs (DETYA). These were:

- Metadata Tools Project, which was concerned with providing a set of documented tools for the management of metadata in the EdNA context. This aimed to:
  - Evaluate the adequacy of the existing tools;
  - Modify the current tools or develop new tools;
  - Provide online documentation to assist promulgation to all education stakeholders;
  - Distribute and support the initial toolset;
  - Provide limited development of the tools to the extent necessary for initial implementation.
- Metadata Development Project, which aimed to provide a retrospective metadata generation toolset to assist all stakeholder institutions to retrofit metadata tags to already existing educational content and collections; and a
- Directory Services Thesaurus designed to improve the search effectiveness and efficiency of the EdNA search engine by developing and implementing an education thesaurus to broaden search patterns aided by keyword.

Challenges

From the new Project Manager's perspective, these projects presented a number of challenges.

Firstly, Metadata is a relatively new field and the proposed tools would be breaking new ground in information access. It was important to define the desired outcomes while at the same time, be open to improvements that might evolve out of ongoing discussions.

Secondly, the tools would have to be available and relevant to the needs of users from all sectors of education from all states. While there was anecdotal evidence of the nature of many of these needs, there was no firm evidence of specific needs of each sector.

Thirdly, the rapid changes in metadata tools and standards internationally would need to be constantly monitored to ensure that the EdNA tools remained viable and able to be readily updated to these changing standards. Finally, strategic linkages to other standards in Australia and elsewhere needed to be considered and the potential for linking to evolving standards would need to be explored. These international developments include IMS [HREF7] in the United States and ARIADNE [HREF8] and the IMesh Toolkit [HREF9] in Europe.

In such a dynamic and challenging environment, it was imperative to select a software developer that had a strong track record in metadata development, that was in tune with international developments and that could work closely with stakeholders through the EdNA Project Manager. He, in turn, recognising a lack of experience in the area, decided that a locally based project manager experienced in software engineering projects would be valuable in setting out the project and technical briefs. As the projects were also government funded, this new sub-manager could also act as an honest broker in selecting the developer and in ensuring transparency in contractual arrangements.

As software developer, the Co-operative Research Centre (CRC) for Enterprise Distributed Systems Technology (DSTC), was selected on the basis of earlier work done particularly in regard to the Metaweb Project [HREF10] on behalf of the National Library and several universities.

DSTC is a joint venture supported by the Australian Government's Cooperative Research Centres Program and over 24 participant...
organisations developing the technical infrastructure for tomorrow’s enterprise.

An initial meeting was held with Dr. Renato Iannella, the then leader of DSTC’s Resources Discovery Unit, which laid the basis for a number of decisions affecting the deliverables and outcomes of the projects.

It was recommended that the three projects be developed in tandem because of their strong linkages. A suggestion was also made that the metadata management tool could be presented as a spreadsheet to perform editing for an entire site, using drag-and-drop between columns with the ability to extract and insert data based on locally stored defaults. It was agreed to develop Statements of Work locally and have DSTC clarify technical specifications. DSTC also agreed to survey current international developments to ensure the projects applied best practice.

Technical Considerations

Stakeholder requirements presented the developers with several technical challenges:

- The tools needed to be able to be run on multiple platforms and with a range of operating systems;
- The tools needed to be able to be locally updated for future versions of the EdNA Metadata standard;
- The tools needed to read and write to different metadata formats, eg. HTML or RDF.

As EdNA stakeholders operate in a highly distributed computing environment that includes PCs, Macintosh Computers (Mac), NT Servers and Unix Servers, any software developed needed to work in all these environments.

For DSTC, this raised the first significant technical issue. Should the software be developed using a language robust to different computing platforms such as Java, or should a native application be used.

For a single platform, using a native application is a logical solution. Its compatibility with the host platform allows it to access all its features yet be both compact and fast. However, a separate application would need to be written for each platform making it an expensive option. Extensive recoding would also be needed when upgrading.

Java on the other hand can be used across multiple platforms, provided a platform-specific runtime Java Virtual Machine (JVM) is available on each client machine. This acts as a converter to the requirements of the particular platform. Compatible JVMs are also present in most current browsers.

Java can be used as applets or as a Java application. Applets need to be downloaded while the software is being run. Users do not have to install software and, because of the way they are handled by the browser, applets have a high level of security. However downloading an applet can take a significant amount of time determined by the size of the applet and the nature of the connection. The user needs to have an active Internet connection and even with a compatible JVM, there is no guarantee that the Java bytecodes will perform identically.

Most of these issues are common to Java applications which also have software installation requirements. However there are fewer security restrictions which allow the programs to make necessary interactions with the operating system.

In summary, if a single platform was to be targeted, then a native solution is much more practical than a Java solution. However, if a cross platform solution is important, then the costs need to be weighed and alternative options explored.

Based on the needs of the EdNA Online community, i.e. for flexibility in a multi platform environment, it was decided to develop the first prototypes as Java applications.

First Prototypes

For the Metadata Development Project (concerned with retrospective metadata creation), DSTC took their prototype Reggie and rewrote it as a Java Application, which must be downloaded and installed on the user’s desktop, i.e. either a PC or Mac. This prototype became known as Eddie.

Eddie is a metadata editor specifically for EdNA metadata records. It provides a Graphical User Interface (GUI) to allow users to create, view and edit metadata records for one resource at a time, and to generate metadata tags for embedding in the resource.

In order to assist users to select values for various elements, Eddie is designed to interface to a number of plug-in thesauri.

For the Metadata Tools Project (primarily metadata management), DSTC needed to ensure the complex functionality required by EdNA would work and therefore developed a prototype called Emma using the Perl programming language. The Emma prototype was developed for Win 32 platforms only, as a set of small programs which were command line driven.

Emma is an application for resource publishers to manage the metadata associated with the resources they publish. It highlights errors and inconsistencies in the metadata records, and provides facilities for making bulk changes to many records at a time. It also enables publishers to gather statistical information about the metadata in their resources.

For the Directory Services Thesaurus Project, the initial focus was on defining the format for an EdNA thesaurus, with the longer-term aim of providing a tool for constructing and viewing the thesaurus, and to integrate the thesaurus into Eddie and the EdNA search engine. A format was agreed upon, but no prototype software was developed.

Following a review of the initial testing and discussion of useability issues by EdNA and DSTC, it was apparent that useability would be considerably improved if all three tools were well integrated and displayed a user interface consistent with the EdNA Online website.
It was decided that EdNA users should have access to three distinct tools, but these tools would in fact be different facets of the one application, *The EdNA Metadata Toolset*.

**Integration**

Integration of the tools was demonstrated in a potential scenario.

While the above scenario provided a coherent depiction of user functionality required for an integrated toolset, it was still a theoretical perspective. Issues relating to software design and user interface still needed to be resolved.

This led to an alternative model, where a single viewing platform performs basic functions and provides options for the appropriate use of *Eddie* and *Emma*. This viewing platform must be able to work with both local and remote resources.

This model can be represented as shown.

This model became the basis for *The EdNA Metadata Toolset*.

**The EdNA Metadata Toolset**

*The EdNA Metadata Toolset* consists of three platform-independent modules:

- The EdNA Metadata Manager
- The EdNA Metadata Editor
- The EdNA Thesaurus Viewer

These modules are highly integrated, in the sense that each will automatically invoke the others whenever their functionality is required. However, as few users will require the complete range of functionality of all modules, they will also be able to invoke each module separately.

**The EdNA Metadata Manager**

This module provides the viewing platform capability that enables users to work with single or multiple resources. It has two modes, which provide either a Resource or Element view of the metadata. Users can switch between the modes by selecting the Resources or Elements tabs in the window.
In Resource mode, it provides users with the facility to build a list of resources, to select one of these and to view its metadata with the option of editing or creating metadata using the \textit{Eddie} module. An icon next to each resource indicates whether its metadata is embedded or detached.

In Element mode, it allows the user to list the metadata elements in one or more of the listed resources and then view occurrences of a single element across these resources. These are shown on a grid that allows the user to select particular resources. The grid also shows possible Language and Schema qualifiers for the element. The user is then able to create or edit element names and values across the resources in a separate Element Editor.

The Metadata Editor

The Editor can be accessed from the Manager in Resource view or separately. The module is based on the functionality developed in \textit{Eddie} and assists users in creating or editing metadata for individual resources. The Eddie screen displays a default minimum set of metadata elements with the available metadata values. These mandatory elements include Identifier, Title, Description and Subject.

Where EdNA Standard metadata is not present, values for these elements will be obtained according to the following rules:

- \textbf{DC.Identifier}: The resource URL (where available)
- \textbf{DC.Title}: Contents of the \textit{<title>} tag.
- \textbf{DC.Description}: Abstract if available, otherwise first 50 words of the text.
- \textbf{DC.Subject}: Keywords from text obtained by a simple word frequency algorithm.
- Other default values can also be provided including Language, Format and Date.

The EdNA Thesaurus Viewer

The Thesaurus module enables keywords selected from the resource to be mapped to a selected thesaurus and then inserted into the DC.Subject element as controlled vocabulary. The thesaurus can be accessed from either the Edit Element screen or from the Resource Editor.

Users can access this module in two ways:

1. It can be invoked directly from the Editor or Manager whenever it is required.
2. It will also be available as a separate application that can be invoked from an icon for viewing purposes.

This module permits users to view data from a number of different thesauri. Data for different thesauri may reside on different servers, but never on users’ local machines. This module provides the means to access and view this data.
There are two basic mechanisms by which users can control the thesaurus to be used:

1. They may choose an education sector within which they are working from a list of available sectors. All modules will provide this facility. Each sector will be potentially mapped to a default thesaurus for that sector.
2. They may select a thesaurus directly from a list of available thesauri. All three modules of the toolset will provide this facility.

DSTC engaged ICE Media to develop The EdNA Metadata Toolset as ICE Media are the software developers for EdNA Online and engaging them would allow already developed EdNA code to be re-used.

ICE Media's familiarity with the EdNA environment also helped to ensure that The EdNA Metadata Toolset would seamlessly integrate into the EdNA Online environment.

Testbedding, training, feedback and review

To ensure general acceptance of The EdNA Metadata Toolset, it was decided to invite representative organisations from all sectors of Australian education (VET, Schools and Higher Education) to run acceptance tests on the prototype. The testbedding process includes three components: a functional test against the Software Requirements Specification to ensure that the tools worked as required, a usability test to check the intuitiveness of the interface and the usefulness of the online help and finally, a relevance test to confirm that the tools met user needs.

A number of institutions, that were not included in the selected testing group, were also asked to download the prototype for familiarisation testing. In this way, it was anticipated that a wide cross-section of useful feedback would be obtained from this group of testing organizations.

Appropriate training for all user groups was seen as an important component of effective distribution and take-up of the tools. Given the diverse nature and distribution of the potential users of the toolsets, a multi-pronged approach was deemed necessary for maximum benefit. Thus an online help system built into the tools would be supported by a series of flexible delivery units and printed wallcharts while the tools would be officially launched with a roadshow to major centres around Australia at which printed User Guides and promotional material would be made available. These would also be distributed via a general mailout.

With this in mind, a series of online training modules are currently in preparation. Using flexible learning technology, it is anticipated that site creators and managers will be able to access these modules as required to provide them with guided practical experience with the tools.

Usage patterns relating to the online contextual help and the training modules also provide useful feedback concerning training needs and usability concerns. Normal usage for new software suggest that users first download and try the software and then access those components of online help that relate to areas of confusion. An analysis of these usage patterns can identify components that need to be addressed in future iterations of the tools.

Alongside the ongoing monitoring of feedback, a major review is planned in October, 2000, six months after initial release. This will involve an online survey and possibly the use of focus groups to clarify whether the tools have met user needs and expectations and to provide directions for further development.

Some of these areas that have already been identified include the development of further sectoral based thesauri. As each sector currently uses a different printed thesaurus, the flexibility of being able to plug-in the appropriate thesaurus was seen as a useful feature of the tools. The process involves conversion of selected thesauri to XML. Thesauri currently identified for future conversion include the Schools Catalogue Information Service (SCIS) Thesaurus for the Schools sector and the Australian Education Index of Educational Descriptors (AEIED) for the Higher Education sector. While work on converting these fell outside the parameters of the current projects, it would be a useful focus for future projects.

Conclusion

For the Project Manager, running these projects provided several important lessons. Firstly there was recognition of the importance of clearly defining outcomes at the outset. While it is in the nature of projects dealing with cutting-edge technology that anticipated outcomes are likely to change, it is easy to lose momentum and direction unless there is a clear picture of what the project is meant to achieve, even if the means to do so is affected by parallel developments.

Retaining clearly defined goals is made easier if ongoing communication is maintained between the stakeholders and the developers. In an environment where stakeholders are widely spread and working independently from each other, this presents challenges and online communications provide the best solutions to this. Even with regular emails though, it is easy to have misunderstandings and a lack of clarity between Project Manager and developers unless there are regular face-to-face meetings. The fact that the developer and the manager were fifteen hundred kilometers apart made this more crucial.

Ensuring that the tools are used effectively still presents many challenges. The principal one lies in having institutions take responsibility for ensuring that their online resources are metadata enabled to a single standard. Discussing this issue, Thomas and Griffin (1998) state that

"The challenge of persuading Internet information providers to implement a standard may be more difficult than any development issues. To achieve a successful metadata solution, we must discover ways to encourage extensive metadata generation."

There is a recognition that the effort of adding metadata to resources requires both an incentive and adequate resourcing. The authors of this paper contend that this will only come about if the institutions can see real benefits in adding metadata.
However despite the obstacles, the EdNA Metadata Toolset has the potential to bring many benefits to stakeholders. By providing the means to simply and efficiently add relevant metadata that fits the EdNA Online Metadata Standard, it gives the opportunity for collection managers to share their resources with other researchers and teachers.

As more resources are made available so the Internet becomes a true library of quality materials. The Thesaurus tool will also enable searchers to more efficiently find what they are looking for using a variety of means including natural language searching mapped to a controlled vocabulary.

The tools will also provide international advantages and economic benefits, particularly to Higher Education institutions but increasingly to schools and VET institutions. As a principal portal for Australian education, EdNA Online is carefully monitored by overseas institutions and potential students when deciding on educational opportunities in Australia. By making institutional profiles, faculties and courses accessible on EdNA Online, stakeholder institutions can take advantage of a huge potential market while reducing the cost of expensive media advertising and overseas travel.

The plug-in design of The EdNA Metadata Toolset allows for the application of a range metadata schemes including Dublin Core and IMS with minimal changes to the actual software. This is particularly significant in view of the recent announcement by Microsoft of support for IMS in its software (Microsoft, 2000).

The toolset also has applications beyond its original parameters as a means of managing and locating resources on a local intranet or extranet within an institutional or corporate environment. The search engines used in these internal networks are increasingly capable of recognising metadata and many are configured to access that based on Dublin Core (which is the basis of the EdNA Metadata Standard).

In summary, the project personnel and developers were pleased to see that the projects provided an enhanced specification that was completed on schedule and within budget. The EdNA Higher Education sector acknowledges the support of the Department of Education, Training and Youth Affairs (DETYA) in financing these projects. The EdNA Metadata Toolset will be available for downloading from the Higher Education pages of the EdNA Online site <http://www.edna.edu.au> from the middle of May 2000.

References


Hypertext References

HREF2 http://www.educationau.edu.au/
HREF3 http://www.edna.edu.au/
HREF4 http://www.w3.org/TR/rdf-schema/
HREF6 http://www.edna.edu.au/metadata/
HREF7 http://ww.imsproject.org/
HREF8 http://www.ariadne.unil.ch/
HREF10 http://www.dstc.edu.au/Research/Projects/metaweb/

Copyright

Michael Currie, Nicholas Moss, Albert Ip & Iain Morrison © 2000. The authors assign to Southern Cross University and other educational and 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 authors also grant a non-exclusive licence to Southern Cross