

Modeling Information to Support Value-Adding: EdNA Online

JON MASON, GRAHAM ADCOCK, & ALBERT IP

Originally initiated in 1995 through a cooperative arrangement between the Ministers of Education and Training of the Australian Commonwealth, State and Territory governments – and at the time primarily as an infrastructure connectivity project – Education Network Australia (EdNA) has since consolidated as an ongoing national collaboration. It is focused on maximising the benefits of the Internet for education. As well as government interests, it brings together representatives from all the education sectors – schools, vocational education and training (VET), adult and community education (ACE), and higher education. In 1997 its most visible product of this collaborative activity, the EdNA Directory Service, was formally launched as a Website. At this stage, this website was clearly positioned to act as the foremost gateway for online educational information in Australia. The potential for resource discovery which promised a stamp of quality (through careful evaluation by educators) was seen as a tremendous value-add for all involved. In late 1998 the EdNA Directory Service was renamed (EdNA Online 1999) to reflect

the fact that communications, interactive services, and networking were recognised to be equally as important for education as quality online information-based content.

While it is important that as an ongoing collaboration EdNA is understood to be much more than an information gateway – and much more than a website – this article is focused on what stakeholders began to refer to in 1998 as a data model for EdNAOnline. In this discussion, we concentrate on illustrating how value-adding emerged as an important process which not only guaranteed quality content but became identified as a strategic consideration and indeed a feature of the EdNA framework. The data model presented here is not intended to be a static picture, but rather a snapshot of an ongoing evolution.

SYSTEM OVERVIEW

During the early stages of the EdNA project a Website (as a directory service) was established as an important central focus. It was thought that such a directory system would be best served through a browsing paradigm. At that time (1995) none of the current major search engines had achieved any widespread presence and the main global directory was the emerging *Yahoo!* which was predominantly accessed through user browsing. Technical advisers to the EdNA Task Force offered a few software solutions with *Oracle* being chosen, mainly because of its support for the dynamic generation of web pages, a feature that was very much in its infancy at the time. A fundamental part of this process was the use of *Oracle's PL/SQL* to generate pages dynamically from information stored in a database.

JON MASON, EDUCATION.AU LIMITED, DULWICH, AUSTRALIA.
E-MAIL: jmason@educationau.edu.au

GRAHAM ADCOCK, GEA.COM, CANBERRA, ACT AUSTRALIA
E-MAIL: graham@gea.com.au

ALBERT IP, MULTIMEDIA EDUCATION UNIT, THE UNIVERSITY OF MELBOURNE, AUSTRALIA
E-MAIL: a.ip@meu.unimelb.edu.au

The dynamic generation of web pages was also very new at that stage and the possibilities seemed endless. Features such as personalised pages were trialed, with users being able to select different graphical themes and starting points in the category tree. These were supported through use of 'cookies,' which, in Australia at least, suffered considerable misunderstanding and criticism from those concerned about privacy issues. Use of logins through user-IDs and passwords were considered to have greater privacy implications and in the end, the personalising features were dropped. With the current, and more sophisticated, implementations of 'profiles' in office-based software as well as on the Web this is now being re-visited for EdNA online.

It was quickly recognised that the browsing facility needed to be complemented by a searching mechanism. There were several alternatives for this and a combination of the Verity Internet search engine with the *Harvest* freeware robot was selected. The robot was set up to gather full-text information based on the items that had been stored in the database. This information included directions for the robot as to how deep and how wide it should follow hyperlinks.

An index was also created from the information held for each item in the database. This provided a much more targeted searching capability and one which was guaranteed to return only quality information, which had been evaluated and enriched with metadata by education experts.

Once the basic technical mechanisms were in place, there was a necessity to identify suitable web resources and to build the content to a level that would be useful. At first this was done centrally but due to the widespread nature of the user and stakeholder base a concept of "devolved (or distributed) administration" seemed to offer better leverage. There were, after all, potentially thousands of content identifiers. To do this it was necessary to provide a distributed administration system which included security control through user-IDs and passwords.

When this part of the system was developed in early 1996, there was nothing available commercially to support such a security system and so it was built into the system using the *Oracle* database and PL/SQL code. A system of security groups was developed and the "state" was maintained through a code which was passed in successive URLs. This was made possible through the dynamic generation of each page which could include the appropriate code with each URL written into every page.

The distributed administration enabled the identification and inclusion of quality content to be built up to a reasonable level. This was sup-

plemented by a form made available to users to suggest their own items for inclusion. Suggested items were stored in the database in a "pending" state until they could be examined and either rejected or included for public access. An e-mail address was also provided for feedback and suggestions which assisted in building up the level of content.

Another feature included in the initial system was the hosting of *Majordomo* mailing lists. Registered users were able to use the administration system to set up and manage their own mailing lists using a forms interface. This was set up to buffer the user from the complexities of the normal mailing list administration and also to control the features that were allowable within the EdNA-hosted lists. Archival of these lists was provided as an option which was achieved through the EdNA system "subscribing" to these lists and groups. The messages were stored and indexed for searching using the same Verity engine which was supporting the main searching facilities.

Current Implementation

These features have served EdNA well and the database of approved items has been built up to some 8,500 items. These items have spawned the full-text indexing of over 230,000 additional web pages which can be accessed through the search facility.

Considerable work has been undertaken on the development of a comprehensive category tree which is used for browsing and allocation of items to parts of that tree. By and large, this process has been fairly disciplined resulting in a reliable, though not entirely engaging, method of resource discovery.

However, by mid 1998 the initial system architecture was showing its age. The increase in information held and the growing use of the service created additional strains. Like any large scale information technology development, this has made it increasingly more difficult to amend the system and to integrate new features.

Performance has been an issue which has largely been dealt with through increasing the power of the hardware. The mechanism of generating every page through PL/SQL code is now dated and is scheduled to be replaced so that the bulk of the processing is more closely aligned with the Web Server software rather than being executed within the database environment.

The freeware *Harvest robot*, used for gathering full-text searchable information, has not been supported for some time and has caused problems, probably due to the size and complexity of the tasks it was attempting to perform. The use

of the *Majordomo* software for EdNA hosted mailing lists has also shown similar problems.

The original “devolved administration” system has worked well, but it no longer provides the flexibility required due to the base of those who would administer particular parts of the system increasing significantly. One example of this is in the control of the mailing lists, where it is desirable for various levels of education participants (e.g., state bureaucracies, educational institutions, individual educators, and even students) to be able to manage lists and possibly allocate list management rights to others. The *Noticeboard* system implemented in 1997, based on NNTP, likewise hasn’t provided the flexibility required in the management of notices or in the selection and presentation of notices to users.

Considerable effort has gone into the area of metadata usage as this is seen as one of the most important strategies for providing a quality and coherent set of resources that can be easily and usefully accessed by users. The “Dublin Core” standard was chosen as a sound foundation from which to proceed.

The issues described above have now provided the opportunity to step back and examine the architecture and direction of EdNA Online and,

in particular, to develop a data model which will provide a solid basis for its development and beneficial use into the future. The remainder of this article describes the current thinking and implementation surrounding this evolving data model.

THE DATA MODEL

Developed in 1998, the first draft of an EdNA Online data model identifies five types of information: documents; items; search indexes; categories; and, linkages. “Documents” are web-

based resources of potential interest to users in the education community. “Items” can be described as value-added containers which consist of all metadata that EdNA has associated with a particular document, regardless of the origin of that metadata. In this regard, EdNA has been an early adopter of Dublin Core metadata as a basis for its own standard which is implemented as a means for enriching its database of stored items (URLs enriched with metadata after

evaluation). The main difference between these two information types is that documents are the raw resources that are created and maintained outside the EdNA environment (in fact anywhere on the Web) whereas items are the units on EdNA which contain the information about each document.

Searching functions of EdNA Online are, based on “indexes” which contain detailed information about the text within each document. “Categories” support the browse approach to information retrieval and enable navigation to useful material as an alternative to search. “Linkage” is an architectural concept; an attempt to extend “traditional” metadata concepts to include the relationship that documents have to additional, associated information such as review material, or other document’s. To date, this concept currently only has a very basic implementation. The higher education sector, in particular, sees a lot of benefit from developing this concept for providing greater flexibility and malleability into the data structures within the EdNA Online database. These information types support the process of value-adding – an essential component of good customer service and in harmony with the value-adding brought about through collaboration

Documents

Documents have not typically been hosted by EdNA Online, although there is a move toward changing this. With its initial emphasis as a directory service and gateway, EdNA Online largely points to documents that reside on their original servers.

The issue of quality is addressed in a number of ways. With documents residing on other servers, and the Web with a characteristic disposition toward changeable content, the quality of a document is usually beyond EdNA’s control. However, the overall co-ordinating body of EdNA – Education.Au Limited – enters into “service agreements” with “registered partners,” or information providers, who in turn ensure that certain quality standards are adhered to. Ultimate responsibility for the quality of a document lies, of course, with the information provider. Each sector has appointed project officers who attend to evaluating content for the service. Complementing this activity, Education.Au also engages in continued monitoring of the service as well as in a range of other activities such as maintaining overall coherence and currency of content. Such processes work to ensure that content accessed via EdNA Online is of a suitable standard.

These information types support the process of value-adding – an essential component of good customer service and in harmony with the value-adding brought about through collaboration

Items

Items, the identifiers which refer to documents, are the metadata stored in the database. The bulk of items currently on EdNA Online have been inputted manually and are either collected by suggestions from the Website users or routinely sought out by project officers. However, there are many organisations within Australia, which perform similar functions within their own localities and maintain similar information that is of interest to their particular constituents. These organisations have also been manually entering their identified items into the EdNA database, a process which has resulted in "double handling" of the information. The potential efficiencies of an automated process have been recognised for some time and this has led to the implementation of a harvesting project, currently being tested.

Items currently have provision for holding content relating to all the 15 Dublin Core elements as well as a further nine EdNA-defined elements. Of these latter elements, the EDNA.Categories and EDNA.Indexing elements are of particular relevance to this article. If an item contains EDNA.Categories metadata the document pointed to by the item becomes browsable "core" item. EDNA.Indexing values are assigned by administrators as a depth guide for the EdNA robot indexer. These documents are full-text indexed and stored in a separate index to the core items. Items may identify more than one category, thereby linking the referenced document to relevant "information zones."

Categories

The browse directory on EdNA Online displays categories arranged as a tree structure. The linkages for this structure are held in a table in the database which is used to dynamically generate the pages which display each section of the tree as the user drills down. The category tree provides a comprehensive overview of Australian education and enables users to browse for "core" or evaluated information.

Adding a category to an item requires a certain amount of education and training in itself and is non-trivial. Attaching categories to documents places the document in particular information zones. Familiarity with the category tree certainly assists in locating information quickly. However, at the time of writing, stakeholders are engaged in the process of examining different categories and their relationships with a view to de-coupling category functionality from the current interface constraints of a drill-down category tree .

Indexes

Indexes are used as the basis for the search facilities on EdNA Online. This task is currently per-

formed by the Netscape Compass Server which, based on the EDNA.Indexing metadata field, gathers text and metadata from pages referenced by core items and creates a search index which is in a format that can be accessed by the Verity search engine. There are currently two main indexes used to return the results to queries made through the search options: the full-text index for non-core items and an index for the title, keywords, and descriptions of the browsable core-items. The first is maintained through the Compass Server robot process and the second is built through a Verity gateway to the information held in the Oracle database. While this dual approach of maintaining two indexes was important in developing the Website, development work initiated in late 1998 aimed at merging the two indexes in a process which has brought search and browse into closer alignment.

Linkage

Linkages are currently implemented only through the existing constraints of the item table in the database where categories are assigned. However, beyond such obvious architectural structure, the EDNA.Review element (designed to store reviews of documents provided by third parties) also relates closely to this concept. Annotations also operate as a linkage to a document. The current implementation in the form of metadata elements is limited by a uni-directional information flow. The authors believe, however, that there is potential for a new mechanism which provides a bi-directional linkages. While this may be difficult to implement it may also break new ground in enhancing a model for collaboration and community building in education.

THE DATA MODEL

Further refinement of the model previously presented, has now led to another abstraction which classifies all data types into three. In this abstraction these types are named Types 1, 2, and 3 respectively.

Type 1 Data

Resources that end users are interested in and often referred to as "content" can be described as Type 1 data. Such resources may include web pages, software, or dynamic data (such as current

Items may identify more than one category, thereby linking the referenced document to relevant "information zones"

stock values or the weather). In the first model, “documents” describe Type 1 data. However, if an end user is specifically interested in Type 2 and Type 3 data then in a way their interest determines that the data they seek also has some Type 1 status.

Type 2 Data

Type 2 data is defined as data which is derived directly from Type 1 data in order to conserve computation or cognitive load in performing some function on Type 1 data. For EdNA Online, the “items,” “categories,” and “indexes,” are Type 2 data, as is metadata (information about data). These are data derived from Type 1 data (harvested metadata, or created by the search engine as an index). In other terminology, Type 2 data can be

seen as an “asset” of search sites (such as *Yahoo!* or *Alta Vista*).

Type 3 Data

Type 3 data is defined as data that describes relationship(s) between data and therefore cannot be derived from Type 1 data alone. It relates to “linkages” described in the first model. Examples of Type 3 data include: the data that describes the hyperlinks between documents; the grouping relationships between data (the EdNA category tree is itself Type 3 data although the metadata element EDNA.Categories in each resource is not); the usage logs of a proxy server about web pages; or, the popularity rating of a web page among similar pages. Clearly, a single datum of Type 3 is fairly useless – just like a telephone, if only one existed it would be useless. The power of Type 3 data

is the collection.

For EdNA Online, Type 3 data is a key to engineering new flexibility into the information retrieval and resource discovery functions of the Website through the design of user-customised pathways. Conceptually, it also acts as a useful abstraction from which to develop mechanisms to support and promote online collaboration and online educational culture. It is expected that the practical implementation of this can inform the

development of appropriate specifications for the “collaboration” module within the Instructional Management Systems (IMS) Project (IMS, 1999).

From a design point of view the application of Type 3 data enables the dismantling of the rigidity of a hierarchically branched (drill-down) category structure and re-rendering it depending on user entry points. This is achieved through the development of a small master set of categories supported by a controlled vocabulary and the use of metadata. This opens the opportunity for a much tighter integration between the browse and search functions and further allows the definition of new pathways to information which can be saved, exported, and shared as educational objects of value.

The model presented here using “Types” of data bears some resemblance to Tim Berners-Lee’s *Metadata Architecture* (Berners-Lee, 1998). However, we see the need for a model which extends further into interactive services, EdNA Online being very much a test-bed in distributed administration as much as it is in the sharing of aggregated educational resources. Berners-Lee’s definition of “resource” is our Type 1 data. His notion of metadata is close to our Type 2 data. An important point here is that our model is not intended to be constrained by its own attempts to be definitive – our definition is intended to be flexible enough to accommodate other forms of Type 2 data not yet revealed.

For our purposes, Berners-Lee’s notion of “link” can be seen as Type 3 data. In contemplating the possibility of innovative uses of other forms of Type 3 data our model extends beyond links as they exist in hyperlinks themselves. Hyperlinks are certainly one of the most important examples of Type 3 data and in many ways the fundamental technology of the Web. Perhaps it may be just wishful thinking but we believe that creative application of Type 3 data will be a key in propelling EdNA Online into its next stage of development.

Both the data models presented above are currently providing EdNA stakeholders involved with progressing development of services for EdNA Online with a tool for conceptualising a “best-fit” approach to matching the requirements of human networks with available online technologies. The central commonality of the data model(s) with the collaborative framework of EdNA can be seen in the processes of pursuing “added-value.”

CURRENT DEVELOPMENTS

Harvesting

It became well recognised from about mid-1997

Both the data models previously presented are currently providing EdNA stakeholders involved with progressing development of services for EdNA Online with a tool for conceptualising a “best-fit” approach to matching the requirements of human networks with available online technologies

that the EdNA Directory Service (now called EdNA Online) required a better method to collect items and improve the search results. As a consequence, an EdNA Harvesting project was initiated after protracted consultation with stakeholders over 12 months. It aimed to harvest quality resources from registered partners who expose collections of their own resources. This will result in these resources being accessed from the EdNA database for Browse or Search purposes. Browsable status depends on presence of <EDNA.Categories> metadata. In the first instance of this project, Harvest Control Lists determine the resources that can be harvested according to a set of agreed rules determined by representatives of all EdNA stakeholders.

The Harvesting project is viewed as an efficient mechanism for the speedy creation of items directly from documents owned by registered partners or from their own repositories of items which they consider would also be of national relevance through exposure by EdNA Online. It is not, however, seen as a complete replacement of manual activity. This is because both styles of item handling are viewed as value-added processes and there seems to be no good reason to privilege one process over the other.

The implementation of Harvesting involves two distinct phases which depend on the presence of EdNA-compliant metadata in target sites. Phase I is based upon the retrieval of metadata embedded within documents, that are specifically referenced by Harvest Control Lists on each registered partner's Web site, by a software robot. This robot also follows hyperlinks from a nominated starting point. In each case the metadata is extracted from the documents and stored in the EdNA Online database of items according to appropriate rules. This is an automatic update which parallels the steps that are followed when an administrator manually enters or modifies an evaluated item. The rules used to update the database have needed to consider the situations where items have been previously manually entered into the database and where future maintenance is to be handled almost exclusively through the automated process.

Implementation of Phase II of Harvesting commenced in August 1999. In contrast to Phase I, the software is configured for harvesting detached metadata (from other repositories). Given that there has been much effort expended in building other repositories, this was seen to be the most sensible approach, but it does raise additional issues such as handling of the third-party items being submitted from multiple organisations. Consideration is still being given

to the manner in which this could be implemented. The focus of Phase II, however, is upon repositories of metadata describing content published by contributing partners. The contributing partners are requested to "expose" the relevant information from their own repositories in a standard format, in this case, the Resource Description Framework (RDF), (RDF, 1999). It is possible that in a subsequent phase of this project that "linkages" information may be tested.

Pathways

During 1997, the Schools sector commissioned an independent study on interface issues for EdNA. It was conducted by the Hiser Group and was later known as the *Hiser Report*. One of the recommendations of this report is that a special interface that appeals to younger users would be a valuable approach. This project has become known as EchidNA, named after a popular, native Australian animal.

Based on an analysis of the data model, the authors discovered that EchidNA can be treated as an alternate pathway through the browse category structure. Furthermore, while browsing is a powerful mechanism for unsophisticated users to discover information, or to obtain an overview of the information landscape, the browse categories are structured to a predetermined tree structure representing only one possible view of the information. Following the *Hiser Report*, EdNA stakeholders came to the view that there is a need for alternate pathways – that is, a multiplicity of entry points into relevant information. Thus, the Pathways project began initially by examining the categories with a view to improving the driving algorithm behind the scenes.

Essentially, the development of a controlled vocabulary which can produce a leaner collection of "master categories"

is seen as a means from which the required flexibility can be engineered. Thus, for example, a possible master category could be "organisation:" categories such as "higher education," "VET institutes," "schools," "educational professional bodies" would naturally belong to this. Groups of categories would typically be orthogonal to each other (there being no dependencies of meaning between groups).

Through this process each master category

One of the recommendations of this report is that a special interface that appeals to younger users would be a valuable approach

would be implemented as a particular metadata element or, more likely, through qualifications of existing elements. Each of these would then contain values selected from a specific, limited controlled vocabulary. Determining which items are displayed at the leaf node of a particular branch of the category tree is, then, a matter of selecting (probably through the search mechanisms) those which contain the appropriate values in specifically nominated metadata fields. This opens the opportunity for a much tighter integration between the browse and search functions, as they are really are the same process.

This approach also provides significantly increased flexibility in the design of the category tree, as the tree, or paths to particular items in an information zone, can be amended without having to change the metadata associated with each document. This can, in turn, be extended to defining other pathways to information with the possibility of users being able to develop their own pathways and store them on EdNA for later use or for referral to other users.

At the time of writing, a Pathways prototype was being tested by stakeholders. It is configured as an optional alternative to the Search-Browse interface, not as a replacement. In summary, it features: (a) an integrated approach to searching and browsing; (b) an approach to classifying documents which is not according to their "position"

in the rigid and hierarchical category tree; (c) an option for users to change the visual representation of the browsing tree structure without necessarily changing the metadata associated with documents; (d) opportunities for users to custom build and save pathways (category and item resource groupings); (e) a strongly constructivist approach to the user interface; and, if toler-

ant of poorly framed search strategies. This latter feature is achieved by building a visual hierarchical representation of the search path created by a user successively modifying their search target.

Technically speaking, search pathways are created with the assistance of software developed by the Distributed Systems Technology Centre (DSTC) – the HyperIndex Browser (HiB) – which interrogates the metadata of documents

in a collection and then returns terms that occur "near" the target term in the user's query. In addition, where the original target term is a phrase, the HiB offers other terms that appear "near" subsets of the original phrase. In this way, pathways (or linked queries) can be built by refining and/or expanding search strategies in any combination [DSTC 1999].

Terms selected by the HiB are returned in a list ordered by a measure of "relatedness". At any time a user may choose a term from this list and repeat the process or retrieve the actual documents that contain a selected search phrase. Where a term returns an excessive number of documents, the list may be reduced by choosing refinement options. It is likely that given the novel approach to resource discovery that is characteristic of the EdNA Online Pathways project that refinements to the concept and its implementation will proceed for at least the coming twelve months.

The Lego Analogy

The documents described in the data model above can be likened to Lego bricks. EdNA Online users are not interested in how EdNA project officers classify information nor how many information types they deal with – they are only interested in obtaining good, comprehensive, and reliable results when accessing the Website. In a similar way, the Lego model builder's interest in building good models is not concerned with how the Lego manufacturer classified Lego bricks into different colours or shapes.

The value of EdNA Online is not focussed on creating the initial bricks. Rather, as a network, EdNA is concerned with achieving a "best-fit" between the national collaborative framework and the system implementation. This is all about value-adding. One further analogy with Lego here is that EdNA Online is therefore more concerned with creating easy-to-assemble kits for EdNA users to build new models out of. Such an approach builds on the foundations established over the last few years.

CONCLUSION

Throughout the development of projects associated with EdNA Online, value-adding can be seen to be a common feature. Collaboration within the EdNA framework can be seen as the pursuit of mutual benefit. At the same time, value-adding can be viewed as an intrinsic opportunity that exists in any network, online or otherwise. In acting on this recognition through implementing a range of strategic initiatives, EdNA can be seen as developing a framework for lifelong learning support. Coupling the flexi-

At any time a user may choose a term from this list and repeat the process or retrieve the actual documents that contain a selected search phrase

bility of multiple user pathways with concepts such as the linkages is expected to provide a very powerful means of resource discovery.

Achieving these aims will require an ongoing balance between taking advantage of the opportunities provided by an expanding range of enabling technologies and the promotion of facilities which result in the maximum benefit to EdNA users. 🌐

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial contribution of the Commonwealth Department of Education, Training and Youth Affairs in making EdNAOnline possible. The authors also wish

to acknowledge the input into the Pathways project of ICE-Media.

REFERENCES

Berners-Lee, T. (1998). *Metadata Architecture* [Online]. Available: <http://www.w3c.org/DesignIssues/Metadata>

Distributed Systems Technology Centre (1999).[Online]. Available: <http://www.dstc.edu.au>

EdNA Online (1999).[Online]. Available: <http://www.edna.edu.au>

Instructional Management Systems Project (1999).[Online]. Available: <http://www.imsproject.org>

Resource description Framework (1999).[Online]. Available: <http://www.w3.org/RDF/>