Theory and practice of application profile development in Australian education and training

Jon Mason
InterCog Pty Ltd, Australia
jon@intercog.net

Helen Galatis
education.au limited, Australia
hgalatis@educationau.edu.au

Abstract

This paper discusses the theory and practice of metadata application profile development in Australian education and training within the context of standardisation activities. A range of historical and theoretical perspectives are discussed, with emphasis upon the fact that there exists scope for a variety of interpretations as to how best to develop and/or express an application profile. Through providing summaries of a selection of Australian initiatives informed by international standards development that span some ten years, it is demonstrated that developing clarity about business requirements can be seen as the most important step in the process. In this broad historical context, it is also highlighted that both theory and practice inform each other. From the diversity in requirements, it is clear that one standard, one schema, or one application profile is not sufficient in the diversity of learning, education, training, and research contexts.

Keywords: metadata; application profile; practice; Australia; edna; EdNA; Vetadata; APSR; ARROW; Learning Federation.

1. Background

Ever since 1997, the Australian Education and Training sector has pursued activities that concern the deployment of metadata in online systems. Motivation for doing so has been largely driven by a perceived need to establish and promote a standardized approach to the cataloguing and management of online resources (Mason & Ward, 2003; Mason, 2004a).

Education Network Australia (originally EdNA, but now edna) first began work on a metadata schema during 1997, with a view to building a directory service of information about the Australian Education and Training sector as well as providing a quality-assured ‘white list’ of Web-based resources for discovery via EdNA Online. The EdNA Metadata Standard was finalised in 1998 as an agreement between EdNA stakeholders (that is, the education sector in Australia), but was never formally ratified as an Australian standard. Also in 1998, the Dublin Core Metadata Initiative (DCMI) established the DC-Education working group for the purpose of identifying the domain-specific requirements of DC-based metadata for education. As is well documented already, this effort provided a ‘template’ for other DC working groups to proceed (Sutton & Mason, 2001).

At this same time, the IMS project got underway, aiming to “define the learning architecture for the Internet” (IMS, 1998). These were the early years in e-learning standards and specifications development, and the IMS metadata specification was its first specification; and it promoted the then draft Learning Object Metadata (LOM) standard that was being developed within the IEEE Learning Technology Standards Committee (IEEE LTSC) (IEEE, 2002).

However, while ‘interoperability’ had become a buzzword within these various communities, the difference in approach to defining metadata schemas represented by the DCMI and IEEE LTSC left a legacy of interoperability problems although the intentions were to do otherwise (Duval et al., 2002).

Despite these obstacles, a number of key initiatives in Australia and New Zealand have helped lay good foundations for building further information-based services useful for the sector. These include:
• The Learning Federation metadata application profile for learning objects (TLF, 2004)
• The Vocational Education and Training Vetadata profile (AFLF, 2005)
• The ARROW Discovery Service for Higher Education (Payne, 2005)
• The New Zealand Metadata Schema for Education (NZ MOE, 2006)
• Various National Library of Australia-based services, such as persistent identification schemas. (NLA, 2006)

However, times not only change perceived needs but ongoing infrastructure development has demonstrated that metadata is critical to facilitating interoperability requirements and good practice in information stewardship (Friesen, Mason & Ward, 2002; Mason & Ward, 2003; Dempsey, 2004; Mason & Galatis, 2005).

By late 2005, key stakeholders, including Standards Australia and the Australian ICT in Education Committee (AICTEC) agreed that there was such a wide range of activity associated with metadata deployment and e-infrastructure development that it was time for a common approach to be re-articulated. Part of this approach would be a requirement to develop clear guidelines on the development and usage of application profiles.

Thus began work on what was initially called a Code of Practice for Metadata Usage in Australian and New Zealand Education and Training. By early 2006, it was agreed to rename the intended outcome to that of a Handbook of Guidelines for Metadata Usage in Australian and New Zealand Education and Training (the Handbook).

The Handbook was scoped to provide practical guidance and support for organisations establishing systems for the creation and management of content in predominantly online contexts concerned with learning, education, training, and research. While online environments are viewed as the primary concern, it is recognised that the Handbook will also be useful for electronic environments more broadly.

Through the provision of use cases, the Handbook intends to show when to choose a particular metadata standard such as the Dublin Core or Learning Object Metadata; when to use Sharable Content Object Reference Model (SCORM) content packaging (ADL, 2004); or when to use or develop a profile such as DC-Education. Its key message is that one standard, one schema, or one application profile is not sufficient in the diversity of learning, education, training, and research contexts. The Handbook is due for publication in August 2007.

Overall, the practice of application profile development in the Australian education and training sector demonstrates a pragmatic approach to the description and management of educational resources that preferences community needs while maintaining a level of interoperability.

2. Theory – Definitions and Purpose

It is unclear how long the term “application profile” has been in use, although it has received widespread adoption and acceptance within the Dublin Core Metadata Initiative (DCMI) communities ever since a paper published by Rachel Heery and Manjula Patel (2000):

We define application profiles as schemas which consist of data elements drawn from one or more namespaces, combined together by implementors, and optimised for a particular local application.

This seemed to be clear enough – in theory. However, at this time, a number of education communities worldwide had already embarked on coming to grips with the implications of the discourse around “learning objects”. Not only had this discourse brought with it an alternative schema for describing learning content (IEEE LOM) – which many stakeholders didn’t understand – but also a range of new considerations concerning the interoperability of e-learning delivery systems. Thus, it seemed that the logical place to begin in creating a meaningful profile was to draw elements from Dublin Core and IEEE LOM. This is precisely what the DC-
Education Working Group proposed in 2000 (Sutton & Mason, 2001). And this is precisely what adopter communities, such as the Learning Federation in Australia, also did (TLF, 2004). Other initiatives at the time, most notably CanCore, preferred to produce an application profile from just one base schema, the IEEE LOM, by simplifying it (Friesen, Mason & Ward, 2002).

Meanwhile, other international efforts were getting underway in the International Organization for Standardization (ISO) context, with the SC36 (Sub-Committee 36) initiating work on what was then seen as a more internationalized version of IEEE LOM. The proposed standard, Metadata for Learning Resources (MLR), initially grappled with developing an abstract data model that might be a super-set of DC and LOM (SC36, 2002). It also looked closely at leveraging existing ISO work, such as ISO/IEC 11179-1:1999(E) Information technology -- Specification and standardization of data elements - Part 1: Framework for the standardization of data elements (ISO, 1999).

In the ISO context, however, the term “application profile” is not present. In its place are three terms: “profile”, “International Standardized Profile (ISP)” and “International Registered Profile (IRP)”. In ISO/IEC TR 10000-1:1998, Information technology - Framework and taxonomy of international standardized profiles, the concept and purpose of a profile is clearly elaborated:

A set of one or more base standards and/or ISPs, and, where applicable, the identification of chosen classes, conforming subsets, options and parameters of those base standards, or ISPs necessary to accomplish a particular function. (ISO, 1998).

The implication of TR 10000 is that conformance to a profile implies conformance to the base standard – which is also one of the goals of application profile development within DCMI contexts, but not stated as explicitly.

2.1. Harmonization

In the twelve months prior to publication of the IEEE LOM in July 2002, there was an urgent need to address mounting confusion within the education and training communities worldwide as to what schema to use, particularly as Dublin Core and IEEE LOM were perceived as competing solutions (Mason, 2005). Thus, as part of a commitment to collaborate known as the Ottawa Communiqué, representatives from DCMI, IEEE LTSC, IMS Global Learning Consortium, EdNA, and GEM (Gateway to Educational Materials) co-authored a paper focused on the principles and practicalities of metadata application (Duval et al., 2002). In this paper an updated definition for an application profile is proposed:

An application profile is an assemblage of metadata elements selected from one or more metadata schemas and combined in a compound schema. [...] The purpose of an application profile is to adapt or combine existing schemas into a package that is tailored to the functional requirements of a particular application, while retaining interoperability with the original base schemas.

This definition in many ways now represents the de facto understanding of how application profiles should be handled in practice – certainly in the domains of learning, education, and training. However, this has not been the end of the theoretical perspective on the subject, as there has been continued refinement of theoretical positions and a growing recognition that related work undertaken in the ISO context is also of significance.

In the DCMI context and building upon the work of Heery and Patel, Thomas Baker (2003) has helped develop an official DCMI definition in which a number of requirements are now clearly spelled out – for example, an Application Profile (AP) “cannot ‘declare’ new metadata terms and definitions; it only ‘reuses’ terms from existing element sets”.

2.2. Clarifying the Purpose

In a recent update to the Australian and New Zealand Standard for Geographic Information (ANZLIC) profiles can be seen as serving three main purposes:
identifying the base standards that are necessary to accomplish specific functions, such as interoperability;

• providing a consistent approach through documentation and guidance to systems implementation; and

• promoting uniformity in conformance testing (AS/NZS, 2005, p.11).

There are many more specific reasons why profiles are developed, but the principal reason is concerned with meeting **specific requirements** of a community of practice (whether ‘localized’ or not) while retaining **interoperability**. In some cases, for example in the DLF/Aquifer Implementation Guidelines for Shareable MODS Records, the W3CDTF encoding of date-and-time (YYYY-MM-DD) is a profile of the more flexible ISO 8601 standard and is recommended in preference for the simple reason that “W3CDTF ensures a more predictable format for dates” (DLF, 2006, p.31).

With the broader parameters outlined above, communities can set about determining their precise business requirements. Exemplars in this regard are now appearing – such as the CEN Workshop Agreement on **Guidelines and support for building application profiles in e-learning** (CWA, 2006). It is anticipated that the **Handbook of Guidelines for Metadata Usage in Australian and New Zealand Education and Training** will provide a similar function.

### 3. Practice

#### 3.1 EdNA (http://www.edna.edu.au/metadata)

Development of the first national (Australian) approach to an education-specific metadata standard commenced in 1997 under the auspices of Education Network Australia (EdNA) and focused on describing quality educational resources that are discoverable on the Web. The EdNA Metadata Standard was also aimed at providing the ‘glue’ to the information management environment in the development of a comprehensive portal service which came to be known as EdNA Online. This was finalised in 1998 – although it was never ratified formally as an Australian standard through Standards Australia processes all Ministerial representatives signed off on it as fit for its stated purpose.

The EdNA Metadata Standard (2005) basically represents an extension to simple Dublin Core and includes elements and vocabularies that are commonly used throughout Australian education and training. It was significant at the time that the ‘Audience’ element – used for the Australian Government Locator Service (AGLS) metadata standard – was seen as possibly one of the most important elements in describing resources useful for learning and education. Practice has shown that this has been true only for very specific contexts.

Experience has shown that institutions and educational organisations, even early on in the adoption of metadata for describing Web-based resources, have opted to create their own application profiles which included elements from the EdNA Standard. Thus, while the Standard was developed with the intention of meeting the needs of all sectors and wide usage was anticipated, the EdNA Standard has been most successful in describing and managing EdNA Online’s (now referred to as ‘edna’) collection of resources and in informing the development of other localised profiles.

Edna collections have been developed using a combination of models – central and distributed. A core team of Information Officers search for and evaluate online resources for inclusion in the edna database on a daily basis. The edna distributed management system also assists in gathering resources from state education departments and other networks. Scheduled harvesting runs have also been successful in rapidly increasing the size of edna’s collections.

For further information, see Sutton and Mason (2001) and Mason (2004b).
3.2. Vetadata (http://e-standards.flexiblelearning.net.au/vetadata/index.htm)

The Vocational Education and Training (VET) Metadata Application Profile (Vetadata) was collaboratively developed by members of the VET community with knowledge and interest in metadata and extensive knowledge of the VET sector. Launched in early 2005, it was developed out of a business need that required the development of learning objects to be used for training purposes. Therefore, from the beginning there were certain requirements that needed to be met by the type of metadata schema developed. For example, learning objects were being developed according to IMS Content Packaging specifications and needed to be compliant with the SCORM.

There were of course other business requirements to be considered, such as the overall framework within which resources could be described and managed by the VET sector. Underpinning this framework was, and still is, the need to foster the usage of interoperable standards for e-learning more broadly.

IEEE LOM v1.0 is the base schema of Vetadata. The complete profile consists of 37 elements, where 26 of these elements require data to be entered. This sub-set of LOM v1.0 elements was selected to simplify the implementation of metadata and was selected based upon utility for resource discovery and sharing of resources across disparate repositories. To meet discovery requirements, a minimum set of elements were identified as mandatory. These include the identifier, title, description and keywords. Where possible, provision was made to facilitate the automation of metadata – for example, because the profile was developed for VET sector usage, it was easy to incorporate LOM element ‘5.6 Context’ with a common term, thus enabling discovery by anyone external to the sector.

3.3. The Learning Federation Metadata Application Profile (http://www.thelearningfederation.edu.au/tlf2/)

The Learning Federation Metadata (TLF) Application Profile was initially developed in 2002 and supports the description and management of TLF content, i.e. online curriculum materials for Australian and New Zealand schools. It references elements from the Dublin Core Metadata Element Set, v1.1, Dublin Core Qualifiers, EdNA Metadata Standard, v1.1, and the IEEE Standard for Learning Object Metadata (IEEE 1484.12.1-2002) (LOM v1.0). It also introduced some new elements, that is, elements not in existing namespace, in order to accommodate local requirements.

Characteristics of the profile include the grouping of elements into five categories:

- management: information that describes resource management
- technical: technical requirements
- educational: educational and pedagogical characteristics
- rights: intellectual property rights and conditions of usage
- accessibility: accessibility characteristics.

For further discussion, see Mason and Ward (2003).

3.4. ARROW (http://www.arrow.edu.au/)

The ARROW (Australian Research Repositories Online to the World) project was initially established in 2003 and funded by the Australian Department of Education, Science and Training under funds allocated to systemic infrastructure. ARROW set out to identify and test best solutions for development of university institutional repositories holding scholarly outputs. A consortium involving a number of Australian universities, together with the National Library of Australia, was established for the project. With a focus on utilising some common infrastructure to “deposit, share, and find”, ARROW repositories manage e-prints (i.e. electronic versions of working documents), electronic publishing, journal papers, and digital theses. Metadata from the current participating repositories is being harvested using the Open Archives Initiative Protocol
for Metadata Harvesting (OAI-PMH) to enable a single search discovery mechanism which is managed by the National Library.

In its initial stages, the ARROW project explored the possibility of using a single metadata schema or profile to accommodate the variety of digital objects stored in ARROW repositories. Both Dublin Core and the IEEE Learning Object Metadata (LOM) were considered but found to be inadequate. It was also the case that Machine Readable Cataloguing (MARC) records needed to be accommodated and, in time, a number of other schemas, such as Metadata Object Description Schema (MODS) and Encoded Archival Description (EAD). It was therefore decided that all metadata that was created at the institutional level would be preserved in deployment of the new service. In order to deliver on this, ARROW has worked with Online Computer Library Center (OCLC) in order to “enhance the OCLC metadata interoperability core (OMIC)” (Payne, 2005). The OMIC provides a means whereby transformations between a number of disjunct metadata schemas is enabled. Dublin Core is used as the ‘glue’ for the resource discovery service across consortium partner repositories.

Within limits, repository managers have been given flexibility to accommodate local needs whilst also meeting national project requirements. To ensure metadata found on a third party record was not lost in transformation, a fluid core schema was developed. If a schema has an element not found in the core, that element was added to the core. This strategy has allowed ARROW to proceed without having to anticipate all types of digital objects and their associated metadata needs from the outset of the project.

The set of principles and rules that accompanied the recommendation of elements included the following:

- The metadata schema needs to be cost-effective
- Metadata creation is a commitment to quality
- For metadata to serve the purpose of future resource discovery, enhanced metadata is required.

3.5. APSR (http://www.apsr.edu.au/)

The Australian Partnership for Sustainable Repositories (APSR) is also a project funded by the Australian Department of Education, Science and Training under funds allocated to systemic infrastructure. It aims to establish a centre of excellence for the management of scholarly assets in digital format, and has an overall focus on the critical issues of the access, continuity, and sustainability of digital collections. A key business requirement has been preservation and archival of resources, and therefore the outputs of the PREMIS (PREservation Metadata: Implementation Strategies) project have been important.

Through its research and development activities, APSR provides guidance and a framework for interoperability to higher education institutions. Its research and recommendations impact on the selection, application and development of interoperable systems by universities and university initiatives.

Since 2006, APSR has been working with the National Library of Australia (NLA) on the development of an application profile of METS (Metadata Encoding Transmission Standard) as part of a broader repository interoperability framework (RIFF). The objectives of the METS Profile Development Project are “to develop and document a core generic METS profile for submission and exchange of digital objects; and METS sub-profiles for each of the APSR RIFF Workflow projects” (APSR, 2007a). The work is also informing development of other NLA work such as its Newspapers Digitisation Project.

As of April 2007, APSR had developed two candidate models for its profiles and not yet completed specification of the profile. However, it has identified some important guiding principles for the profile (APSR 2007b). It would:

- be registered with the Library of Congress
• be implementation and content agnostic
• set basic rules for elements and attributes
• identify common schemas, vocabularies, and structures
• be based upon existing NLA work.


The Carrick Institute was launched in August 2004 as The Carrick Institute for Learning and Teaching in Higher Education, and was established to:

• Promote and support strategic change in higher education institutions for the enhancement of learning and teaching, including curriculum development and assessment;
• Foster and acknowledge excellent teaching in higher education;
• Develop effective mechanisms for the identification, development, dissemination and embedding of good individual and institutional practice in learning and teaching in Australian higher education;
• Develop and support reciprocal national and international arrangements for the purpose of sharing and benchmarking learning and teaching processes; and
• Identify learning and teaching issues that impact on the Australian higher education system and facilitate national approaches to address these and other emerging issues.

Currently, work is underway to develop an online environment to support online collaboration, sharing and exchange of resources and knowledge, and to foster the building of networks and relationships amongst higher education practitioners and researchers. It is envisaged that the infrastructure required will include discovery of learning and teaching resources, address curriculum issues of interest to the sector, support pedagogical aspects of resource development, and provide information and services that promote scholarship and leadership. Web 2.0 technologies will be deployed to encourage and foster user input and commentary, and as such, metadata requirements are being investigated to implement the various requirements. It is currently proposed that a set of schemas will be used to align as closely as possible to the business requirements of this project. For example, a metadata schema will be developed to capture information about people and organisations. Other schemas will be considered for resources and services. Another way of describing this system is as a super schema which consists of sub-schemas. The connections between the schemas will be usage of common vocabularies which will link people to resources and resources to people.

In the process of determining the metadata requirements for the proposed service, it has therefore become clear that any finalised metadata profile will need to accommodate a number of schemas. Moreover, if ‘Web 2.0’ approaches are to be harnessed, then there will also be provision for schema-less metadata!

4. Discussion

4.1. Cascading Models

Practices indicate that developers or profiles within an organisational context tend to adopt and customise existing profiles rather than start from a preferred base schema. An Australian example of this is where the New South Wales Department of Education and Training adopted Vetadata as its base schema. The decision was based on business objectives and the need to be interoperable with the VET sector, that is, there was a need to achieve a high level interoperability. In practical terms, it means that there is not only a common schema but also common usage of vocabularies to describe the elements of the schema. Thus at this level there is interoperability at the sector level but also at the standard level, as Vetadata shares IEEE LOM v1.0 as its base schema.
In this example, other requirements, particular school sector descriptors, necessitated the need to build on the base schema and borrow elements from other schemas such as The Learning Federation Schema. In this instance, a two prong cascading effect is created where the local schema needs to be interoperable with the preferred school sector schema as well. Even though a narrow view of the world is reflected by the localized schema, we find that a high level of interoperability is also achieved at the sector level. This example illustrates how interoperability can be achieved at a range of levels. However, the opposite can also occur when a customized profile is adopted as the base schema, that is, the narrowing or localized adoption can reduce interoperability while striving to maintain local relevance.

In short, we are, and perhaps more so in the future, experiencing a situation where profiles tend to be built on elements borrowed from not just one but two or more base schemas. An example of this is The Learning Federation schema. Such practices create complexities and mapping difficulties which are inherent in particular schemas.

4.2. Profiles More Broadly in Learning, Education, Training, and Research

In very general terms, there are two primary communities that have informed the theory and practice of application development within the Australian education sector: e-learning standards development and digital library research and development. While historically distinct, these communities share some common interests such as the management and interoperability of Web-based information resources (Duval et al., 2002; Mason, 2005).

Much of the activity in application profile development in Australia has been informed by a number of international activities spanning both these communities, many of which are currently active or yet to finalise their outputs. In particular, the harmonisation effort underway as a result of the DCMI-IEEE LOM Taskforce is being watched closely by many stakeholders active within the projects discussed above.

Looking more broadly across the learning, education, training, and research domains, there can be seen many examples of profiles of standards and standardised profiles which are likewise informing and influencing Australian practice to implementing metadata. For standardised approaches to delivering e-learning, the stand-out example is SCORM, which profiles a number of standards and specifications such as the LOM.

In terms of providing insight into the broad requirements of the education sector globally, the CEN/ISSS Application Profile Registry provides information about application profiles which have been developed for use in educational contexts. Specifically, it aims “to create a freely accessible centralised site to find and add application profile descriptions” (APR, 2007).

The re-activation and re-focusing of the DC-Education community on developing a modular profile is clearly in step with other international activities. Achieving signoff on this in the short-term will be of enormous benefit to the wider adopter community and provide some useful direction in application profile development.

Within the formal ISO context and also relevant is the fact that the SC36 activity around metadata has not been an easy exercise in consensus building. But this has partly been due to difficulties in separating out the functions of a standard and an application profile. This has now been resolved with a flexible and modular approach being pursued (SC36, 2007).

5. Conclusion

Our exploration of theory and practice in the broad context of ongoing standardisation efforts, if nothing else, illustrates that the development of metadata application profiles has been an evolving story. There are as many approaches to the creation of profiles as there are application profiles. However, one thing remains constant within the educational contexts, and that is that the base schemas tend to be either DC-based or LOM-based.
Practice has shown that minimum level interoperability is assured through the use of the commonly shared elements of title, author, identifier, description and/or keywords. In a way, this is not all that dissimilar to John Kunze’s “kernel” approach to the “core”: who, what, when, and where (Kunze, 2001). Nevertheless, in educational settings, the key question is whether this is adequate. For some purposes it will be; for many, it will not. In many cases, the usage of common vocabularies will be just as (if not more) important as any particular schema. Thus, it remains that clarity of purpose is fundamental to the choice of a metadata schema and how an application profile is developed. For, as Lorcan Dempsey has said:

In network space, metadata will be associated with everything that moves … supporting multiple operations: multiple types of information objects; collections; services; people; organisations; places; terms; formats; rights (Dempsey, 2004)

References


