

Visualising Interoperability: ARH, Aggregation, Rationalisation and Harmonisation

Michael Currie, Meigan Geileskey, Liddy Nevile, Richard Woodman
mfcurrie@ozemail.com.au, Meigan.Geileskey@dpc.vic.gov.au,
liddy@motile.net, richard.woodman@dpc.vic.gov.au

Abstract

This paper proposes a visualisation of interoperability to assist real-world deployment of metadata.

For some time, resource managers in many organisations have been acting on faith, creating 'standards compliant' metadata with the aim of exposing their resources to provide interoperability in discovery activities. In some cases, their faith has led them to miss the very essence of the work they are doing, and they have not got what they worked for.

The authors report a case study involving government agencies in Victoria, Australia. A number of departmental agencies have implemented, more or less, the DC-based Australian Government Locator Service (AGLS) application profile, at least for their web resources. They have done this with care and precision, with the long-term aim of developing a fully interoperable system. In the case study, typical would-be records for seven government departments were studied and it was shown that the tiniest, and typical, variation in use of the standard can be expected to thwart the aims of interoperability in significant ways.

In the context of the government's move to seeking interoperable metadata for all resources, including those within document management systems, the authors make visible how a small 'creep' can lead away from interoperability and how it might be contained in the future. They use a 3-step approach of 'aggregation, rationalisation and harmonisation' to expose the problems with 'nearly good enough' interoperability and the benefits of good interoperability, and encourage true harmonisation.

Keywords: Metadata, interoperability, aggregation, harmonisation, rationalisation, Dublin Core, government, AGLS.

1. Introduction

This paper aims to describe a method used to demonstrate how even small variations in the interpretation and use of standards can affect interoper-

ability efforts. It proposes a visualisation of interoperability, particularly in order to make it more understandable to less-expert metadata managers. The managers in the authors' context were information managers of government departments, and the aim was to develop metadata that would lead to the discovery of each and every document, or resource, in a government intranet as might be required if a minister were questioned in parliament on a particular topic. The aim of the project was to encourage the managers to strive for increased interoperability. It shows how easily the move to local applicability among the various agencies has led away from interoperability in practice and how it might be contained in the future. To do this, the authors use a 3-step approach of 'aggregation, rationalisation and harmonisation' and expose problems with 'nearly good enough' interoperability.

2. Literature review

While there is a lot of literature on the mechanics of interoperability, there is very little that attempts to make it understandable to non-specialists, or connect it with everyday practice..

From the very start, the ability to search across a range of resources was a high priority. A number of authors have attempted to define and explain the function of interoperability but they generally do so in the context of metadata that is not fully interoperable. They are working on strategies for bringing together collections. Typically, Hunter (2001) states that interoperability is intended to "enable a single search interface across heterogeneous metadata descriptions, to enable the integration or merging of descriptions which are based on complementary but possibly overlapping metadata schemas or standards and to enable different views of the one underlying and complete metadata description, depending on the user's particular interest, perspective or requirements". Arms (2002, p. 3) argues that "the goal of interoperability is to build coherent services for

users, from components that are technically different and managed by different organizations”.

In the case of a single author, a government, working through its many agencies within an intranet, the differences between the collections can be expected to depend upon the different domains of operation. This does not necessarily mean they should be technically different even though variation may occur in levels of granularity of description of content.

The 1995 minimalist standard known as the Dublin Core was designed to meet the basic needs of different communities for “specifying metadata to support cross-domain resource discovery on the Internet” (Weibel, 2000). The Dublin Core Metadata Element Schema (DCMES) operates within the extensive Warwick Framework (Lagoze, 1996) which provides a modular structure to DC metadata to enable it to accept not only metadata from other standards but local qualifiers to existing elements and separate elements to meet the specific needs of the client groups. In order to maintain consistency in adding qualifiers, interoperability strategies involve carefully defined structure and registries to record and make available the different local applications. The composite view of DCMES is endorsed by what are now known as ‘application profiles’ (Heery, 2000). While the use of application profiles provides a solution to ensuring local specificity, current use of these often limits interoperability rather than enhances it, as shown below.

Heery (2000) argues that there is often disparity between the practices of the standards makers and the implementers. The former group views the power of metadata in consistent adherence to the accepted standards. Implementers, on the other hand, need metadata that serves their specific needs. Seeking to produce an effective, differentiated service, they often assume that the restrictions of adhering to accepted standards limit the utility of metadata to their users. The authors investigated this issue and propose that by sticking to standards, in the case of an intranet, the agencies involved could achieve both the power and the utility they seek.

Hunter (2001) argues that “significant new initiatives ... are demanding application profiles which combine elements from a number of different existing standardized metadata schemas whilst maintaining interoperability and satisfying their own specific requirements through refinements, extensions and additions” (p. 1).

Problems of interoperability can lie in meeting the needs of different standards. In highlighting the need to reconcile the competing/complementary needs of DCMES and INDECS/DOI, the rights focussed metadata adopted by the publishing community, Bearman, Rust, Weibel, Miller and Trant (1999) proposed using a common logical model, the IFLA Functional Requirements for the Bibliographic Record (FRBR). “Translating both the INDECS requirements and the DC requirements into the IFLA

model provided the framework of a common logical expression for the two perspectives (in which) common semantics can be identified for each metadata element” (p. 6). This approach, using a third model to promote the interoperability of two others, has been replicated in a number of contexts.

In another approach, Blanchi and Petrone (2001) propose yet another digital architecture for managing and sharing metadata and metadata schema between digital libraries. After describing and identifying metadata schema, using a DTD that specifies the various attributes expressed in XML and the CNRI Handle System for schema identification, they used the DTD to develop a framework geared towards making metadata instances, schema and services into first class network objects. Using CNRI’s Digital Object Architecture, these digital metadata objects were then deposited in data elements and given an Interoperable_Metadata content type regulated through a metadata registry to enable dynamic metadata conversion. As is apparent, the process is complex and there are issues of scalability. The process also requires development of software modules for each schema.

The Open Archives Initiative (OAI) develops standards and protocols for metadata harvesting to achieve interoperability between its data providers. Its use of “unqualified Dublin Core as the common metadata set was based on the belief that the common metadata set in OAI is explicitly purposed for coarse granularity resource discovery. Community-specific description, or metadata specificity, is addressed in the technical framework by support for parallel metadata sets. The technical framework places no limitations on the nature of such parallel sets, other than that the metadata records be structured as XML documents, which have a corresponding XML schema for validation” (Lagoze, 2001). While the OAI approach is useful in some contexts, its use of unqualified DCMES and retention of all optional elements means that the resultant interoperability is surface level and not suitable for the government intranet.

The authors were motivated to help the agencies involved in this case study avoid the difficulties reported to be associated with post-hoc harmonisation.

In dealing with the human and practical problems of people committing to metadata implementation, Arms (2002) adopts the term “levels of interoperability”. He argues efforts to enhance and enforce interoperability can be seen as a balance between the cost of acceptance and functionality. He argues that “if the cost of adopting a standard is high, it will be adopted only by those organisations that truly value the functionality provided” (p. 4).

From these results and their own evidence of user behaviour, the authors argue that unless the process of developing interoperable metadata is simplified

and made clear to collection owners, along with the benefits, there will be problems with take-up despite the technical research. They argue that the current problems in developing interoperability solutions, identified in the literature above, have as their basis the multitude of variations to be found in most metadata records. These variations, in the case study in particular, include local adaptations of standardized metadata, local terminologies and alternative spellings and words, as well as trivial errors of use and of grammar and spelling.

3. Case Study Overview

Although the case study reports work that involved only a few metadata records, it did involve the future of the whole-of-government intranet and all departments of government. The problem was, how could such a wide audience be encouraged to engage with the existing problems in their metadata implementation. They were soon to be involved in extensive metadata creation for all government documents but were already questioning, after a number of years of working with AGLS metadata, the effectiveness and expense of the process. Departmental information managers were involved in the process described.

It was originally assumed that the fundamental problem was not in the process but maybe in the commitment to it. The process described showed that it was indeed the process, but that by reducing the effort and clarifying the process, government data managers could take a more active role in the production of interoperable metadata and so, in turn, achieve improved results in resource discovery and management.

The reported project aimed to achieve the following:

- Find ways to illuminate the current limitations in interoperability resulting from existing metadata practices;
- Articulate the cause of the problem;
- Develop a shared strategy for improving the interoperability, and, as it emerged,
- Encourage data managers to develop a single, comprehensive metadata application profile, derived from the current requirements and foci of all users, that does not place limits on high level local specificity but enables deep and comprehensive metadata interoperability across the particular participant group.

The result has been increased interest in harmonisation of the metadata, and the development of a shared, more detailed application profile (so far, for the six most commonly-used elements).

The on-going project aim is to help government data managers achieve complete and deep interoperability. This may be achieved now through the development of a single application profile based on existing records that incorporate metadata specific to

agencies within a framework that can be accessed by all. Individual agencies might choose to operate with subsets of the application profile, in the knowledge that their application profile is fully harmonised with those of all other participating agencies. In addition, control of vocabularies and formats for metadata values has been recognised as important for interoperability, and this will be increased. The current proposal is for collaborative extension of the original AGLS profile, with greater specificity to suit the needs of the local state government.

4. Making Interoperability Visible - the ARH process

In making interoperability visible, the authors' approach is to *aggregate* all metadata elements from the resource collections, consider the processes that could be used to *rationalise* the aggregated set of elements and then show how the agencies might work together to *harmonise* the resulting application profile. This process is referred to as ARH – HA!: visualise the processes of aggregate, rationalise, and harmonise in order to be motivated to harmonise commonly-owned, distributed, heterogeneous metadata collections.

Step one, the aggregation stage, involves the collection of data, and analysis of element usage and variation. During this stage all collected metadata tags are added to a table or spreadsheet. Any discernible variations in element names, formats or values that could confuse a search engine, such as different spellings and alternative element names and qualifiers, are recorded separately. At this point, all the differences in the use of elements are made visible and it is a simple step to seeing that interoperability could be enhanced by adding qualifiers to increase conformity and define specificity. While this may increase interoperability, it would not lessen the number of element types, or simplify the application profiles in use.

Step two is consideration of the rationalisation of the metadata. This step involves careful examination of the different metadata elements looking particularly for unnecessary variations, such as when the same value is contained in elements with different names (and namespaces) or when the same elements contain different types of values, such as different date formats. This process makes it easy to see the possibility of considerably lessening the number of types of elements, and so simplifying the application profiles and increasing interoperability.

Step three is the harmonization of the metadata. To ensure that metadata operates as a powerful and accurate communications instrument for all resources from all agencies and departments, data managers consider the use of elements and decide on harmonised approaches to their use in order to develop a shared application profile. As they agree on for-

mats or vocabularies, they see the number of elements deployed across the agencies reduced.

Functional success of the three-step visualisation process is measured by whether or not those who participate in the process do commit to harmonising their application profiles, whether it becomes ARH-HA!

5. ARH and Victorian government resources

Victoria is a state in the federation that is Australia. The Government of Victoria was an early adopter of the DC approach. More recently, government agencies have been attempting to improve access to their records and public documents through the use of Australian Government Locator Service application profile (AGLS) DC-based metadata. Government policy states that all Victorian government agencies should use the AGLS application profile to describe web-based online information resources. This is in line with the Federal Government's metadata directive to its departments. To date, the Victorian policy is, however, advisory rather than prescriptive. Agencies have been, to greater and lesser degrees, left to their own devices - to 'go it alone'. In fact, in the absence of any guidance other than the central policy, adoption has been spotty and often confined to what might be described as web 'brochure-ware'. Deep adoption of a unified approach to metadata has been difficult to achieve although it is now required.

In practice, departments and their agencies have used AGLS metadata and customised their application profiles, more by implication than design. Different departments use metadata for different sets of resources ranging from online, web-style public resources (classified as brochure-ware by the authors) to all resources including those embedded deeply within databases and document management

systems, and never intended to be widely accessed. In addition, departments differ in their use of metadata, some seeing it as possibly useful for export to those who may need to know of the department's resources and others using it to drive their internal resource management systems.

Recently, one department has been given responsibility for developing a whole of government intranet and another for developing a whole of government public 'brochure-ware' gateway. Working on the intranet, the authors have been concerned about how to achieve high levels of interoperability of government resources. They developed the ARH activity in the process of tackling their own concerns, conscious that they were also providing a better framework in which the other agency might develop the public gateway.

First, the authors decided to test the interoperability of existing metadata records. This had been done before but it had never led anywhere. Nevertheless, a series of requests were made to each participating agency, starting with a copy of their application profile, then for sample records showing the use of the profile, and finally for a set of metadata records for analysis. This last request was made when the authors decided to experiment with the ARH process to provide a concrete demonstration of interoperability across the different sets of metadata. In all, 29 records were obtained from six of the participating departments.

The first step, aggregation of the metadata was done by creating a spreadsheet of all the records provided in order to determine variations in the metadata. To approximate the requirements of machine based searching, any variations in element names or format were treated as different elements. This was also applied to value strings where these would be interpreted differently.

Figure 1 shows a small section of the resulting spreadsheet demonstrating the kinds of variation that immediately became visible.

Rec no.	DC.Identifier	DC.Title	title	DC.Creator	DC.Creator name Corporate	DC.Publisher	DC.Subject	DC.Keywords	subject/keywords	DC.Description	description	DC.Language	DC.Language	DC.Date created	DC.Date modified
1	http://www.vic.gov.au	Victorian Government home		DPC		State of Victoria	Victoria; Victorian Governm	Victoria; Victorian Governm		The Victorian Governm	The Victorian Governm	en		1999-07-20	2001-03-2
2		Department of Justice		Department of Justice			Justice Consumer Affairs			The Department of			en-au		
3	CD3EEAECE5E98E1CCA256	Marriages (level 2 overview)	Department of Justice -	Department of Justice		Department of Justice	24 hour help	24 hour help					en-au	2002-02-01	
4	8BAFAF7D34AD535	Marriage Certificates	Department of Justice -					certificate, marriage, single						2002-16-02	
5	1000151	Fishtank		Richard Billingham				BETAVILLE, EXEMPLA		A high-rise flat in the U.K.		English Regional accent		1998-01-01	
6			Department of Education		State of Victoria, Department			education department,		Education Victoria is the		en-AU		2001-10-09	2001-12-2

Figure 1. Sample of the Aggregated worksheet

Element name	Examples of values provided
DC.Title	Victorian Government home page Department of Justice Marriages (level 2 overview) Marriage Certificates Fish tank
DC.TITLE	SOFWeb Front Page
DC.title	Victorian Education Channel
title	Department of Justice - Births Deaths and Marriages - Marriages Department of Justice - Births Deaths and Marriages - Marriages - Marriage Certificates Department of Education & Training Victorian Curriculum and Assessment Authority, Australia Arts Matters Copyright, Trade Marks And Disclaimers Victorian Education Channel - Welcome Page

Figure 2, deliberately out of focus as if seen from a great height, shows graphically the 'spottiness' of the metadata with 49 different metadata elements being used (across the top), more and less, and in a wide variety of ways (see below).

This table, displayed in full and used to illustrate the process of searching, was a useful tool. Participants posed search queries and looked at what a machine would discover. From only twenty-nine records, 49 different metatags were generated. These results made visible why a normal search across these records would produce inferior results.

In fact, variations noted within the metadata by participants included:

Element Name Variants

- Inconsistent case: e.g. DC.Title/TITLE/title; EDNA.Userlevel/UserLevel
- Non-standard names: e.g. DC.Keywords
- Non-standard qualifiers: e.g. DC.Description.Abstract
- Non-standard abbreviations: e.g. DC.Lang

Field Selection

- Standardised v non-standardised element names:

e.g. use of 'description' v DC.Description

- Use of created metadata names: e.g. Custodian

Value string Variants

- DCMES suggests certain type of value strings be used for each element/qualifier, to assist search engines
- DC.Identifier: URI recommended, other identification numbers given without qualifiers
- DC.Date: Recommended ISO8601 standard uses yyyy-mm-dd. Other formats used include yyyy, yyyy/m/d, yyyy-dd-mm
- DC.Format: Controlled vocabulary recommended a) Non-standard terms used e.g. VHS (PAL) b) Incorrect case e.g. text/HTML
- DC.Language: DC recommends RFC1766. Variants include en, en-au, en-AU
- DC.Type: controlled vocab recommended. Non-standard Types used e.g. references and materials
- EDNA.Version: reserved for version of EdNA Metadata Scheme
- Qualifiers embedded in values: e.g. DC.Publisher CONTENT="corporateName=State..." v DC.Publisher.nameCorporate=

Figure 2.

- Non-standard proper names e.g. DPC v Department of Premier and Cabinet
- Generally inconsistent use of capitalisation and punctuation

Overall

- Most element variants due to non-standardised application of capitals, punctuation, spelling
- Users seem unaware of Application Profiles: Little use of collection specific qualifiers to enhance specificity

The results were then *rationalised*. The authors considered ways to reduce the size of the resulting metadata set without detracting from functionality. Elements distinguished only by grammatical errors and variations or spelling can be merged; elements not used across the different collections can be removed without loss, while those with variations of the same name (such as abbreviations of element names) can be merged. For instance, DC.Title, DC.TITLE and DC.title can all become DC.Title in HTML or dc.title in RDF. This reduces the original 49 separate tags to 42.

The final *harmonisation* process is a more complex task as it involves identifying opportunities to make decisions about best practice in the use of metatags. The first step is to look for chances to merge non-standard elements into the closest related standardised ones, e.g. merging DC.Keywords into DC.Subject. This can also apply to the use of non-standard qualifiers. Search tools, designed to access existing variations of the metadata across the various collections, can then use the merging and mapping processes identified at this stage.

Significant variations in the Victorian government metadata records were found in the format and selection of value strings and content. In describing application profiles, Heery and Patel (2000) state that application profiles may “specify permitted schemes and values” such as a particular controlled vocabulary or item string format. Thus the application profile might specify a format to be used for DC.Source instead of free text. In harmonising the records provided, each element value string needs to be addressed separately. The precise form and detail of each needs to be addressed by focus groups to incorporate the necessary input and ownership of the different stakeholders. In some cases, however, there are established practices for DCMES that can be followed more precisely, such as the use of ISO8601 for date fields.

The resulting metadata set, or harmonised application profile should then allow the specific detail of individual metadata collections to be shared and accessed by other departments and users. This could be achieved by providing clear information about the application profile in a shared registry. Given such a registry, in the future individual agencies could select from established elements and qualifications or contribute finer grained qualifiers of use without loss of

interoperability. (The department representatives have now indicated their interest in establishing such a registry).

6. Detail of Case Study Methodology

Size of the sample: Seven departments were originally contacted with a request for ten records each within only three weeks. In fact, initially only five departments replied and only two of these provided 10 items. By the time that results were compiled, six departments had supplied a total of 29 records. While this was not a statistically significant sample compared with the number of metadata records owned by the departments, it proved sufficient to provide a demonstration of the process.

Quality of the sample: The departments were given few guidelines on what to provide for the activity. The request was simply to send examples of the documents and associated metadata from their website or intranet. While one department sent a broad range of documents representing different sections of the department, others sent information sheets or technical papers. Most records described ‘brochureware’ and it was noted that the associated metadata was fairly brief.

Quality of the metadata: As mentioned, there was a wide diversity between different documents and their associated metadata. Of the twenty nine records, 11 could be regarded as comprehensive (ie with at least 9 separate metatags), 3 were very brief (fewer than 4 tags) while 15 were between these. It was interesting to note that even in such a small sample, a wide diversity of styles was apparent.

The rationalization process: Selection of the criteria for rationalization caused some discussion among the participants. While those used seemed to be logical, it was agreed that there was an element of subjectivity involved and these criteria might vary based on the particular samples provided.

Harmonization: The harmonisation process has not yet been completed. It involves focus groups of collection owners meeting to agree on appropriate and useful metadata based on their specific needs balanced against the aims of interoperability. What is important is that the departments have agreed, after participating in the process, to work together on this harmonisation process. In one sense, it is as if the metadata process is being started afresh. This is not the case. Participants who have large collections of metadata are meeting to iron out wrinkles that have developed over time, and this activity is able to draw on five years’ experience with metadata creation and use. It is better-supported by this experience than was the first attempt, and it comes at a time when a powerful outcome motivates it. Whole-of-government interoperability is no longer expected to be achieved by letting agencies work independently and hoping that technologies can be developed to reintegrate the

metadata post-hoc. The need for planned interoperability has become visible and is now being made operational.

7. Conclusion

At a meeting of information managers from the government departments working together on the intranet, collection owners expressed satisfaction with the results of the work done so far. One stated that the display (visualization) of the interoperability of the current metadata had made her see the importance of standardization. Another said that from now on his agency would increase its efforts to generate more useful metadata for their collection.

The final application profile has yet to be delivered but it is anticipated that it will be accepted much more readily than previous profiles because of the local input in developing it. The authors conclude that this process would not have been undertaken in the context if the ARH process had not been developed and attracted the managers' participation. Particularly as it was not the first attempt to achieve the outcome, but was successful. Further, the authors recommend the activity as being useful to those working with information managers and others who are developing practices and implementing established application profiles. The visualisation of interoperability seems to be useful in such a context.

References

Arms, W et al. 2002. 'A Spectrum of Interoperability: The Site for Science Prototype for the NSDL'. D-Lib Magazine 8(1) Jan 2002. [Online] <http://www.dlib.org/dlib/january02/arms/01arms.html> [Accessed 2002-06-12].

Bearman, D. et al. 1999. 'A Common Model to Support Interoperable Metadata'. D-Lib Magazine, Jan 1999 [Online] <http://www.dlib.org/dlib/january99/bearman/01bearman.html> [Accessed 2002-06-13].

Blanchi, C. & Petrone, J. 2001. 'Distributed Interoperable Metadata Registry'. D-Lib Magazine 7(12) Dec 2001. [Online] <http://sunsite.anu.edu.au/mirrors/dlib/dlib/december01/blanchi/12blanchi.html> [Accessed 2002-06-10].

Heery, R., Patel, M. 2000. 'Application Profiles: Mixing And Matching Metadata Schemas', Ariadne Issue 25, Sept 2000 [Online] <http://www.ariadne.ac.uk/issue25/app-profiles/> [Accessed 2002-06-13].

Hunter, J. 2001. 'MetaNet - A Metadata Term Thesaurus to Enable Semantic Interoperability between Metadata Domains'. JoDI 1(8) February 2001 [Online] <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/> [Accessed 2002-06-16].

Lagoze, C. 1996. 'The Warwick Framework: A Container Architecture for Diverse Sets of Metadata', D-Lib Magazine, July/August. [Online] <http://www.dlib.org/dlib/july96/lagoze/07lagoze.html> [Accessed 2002-06-16].

Lagoze, C. & Van de Sompel, H. 2001. 'The Open Archives Initiative: Building a low-barrier interoperability framework'. [Online] <http://www.openarchives.org/documents/oai.pdf> [Accessed 2002-06-19]

Weibel, S., Koch, T. 2000. 'The Dublin Core Metadata Initiative'. D-Lib Magazine 6(12) Dec 2000. [Online] <http://www.dlib.org/dlib/december00/weibel/12weibel.html> [Accessed 2002-06-16].

URIs

Open Archives Initiative. <http://www.openarchives.org/>

The Dublin Core Metadata Initiative. <http://dublincore.org/>