Was CIMI too early? Dublin Core and Museum Information: metadata as cultural heritage data

Liddy Nevile
La Trobe University Australia
Tel: +61 4 1931 2902
Fax: +61 3 9479 3060
liddy@sunrisersearch.org

Sophie Lissonnet
James Cook University Australia
Tel: +61 7 4781 4111
Fax: +61 7 4779 6371
slissonnet@netspace.net.au

Abstract:
In 2000, CIMI reported on the use of Dublin Core metadata in the museum context and proposed a broader set of elements for museums. This paper examines that decision in the light of developments in the DCMES and reaches a different conclusion. The conclusion is based on experience with a virtual museum of Indigenous culture when metadata application profiles and schema cross-walks were available, unlike when CIMI investigated the same issue. In particular, it is now possible to use rich descriptions (data) and Semantic Web technologies and maintain the interoperability of Dublin Core metadata.

Keywords:
Dublin Core, metadata, application profile, localisation, museum, cultural heritage, Quinkan Indigenous community, case study.

1. Introduction

The CIMI Consortium (Consortium for the Computer Interchange of Museum Information) investigated the original simple Dublin Core Metadata Element Set and its application in the museum context. The CIMI experiment was conducted in two phases between 1998 and 2000 (1). Started in mid-1999, Phase 2 of the test bed concluded that the DCMES was too focused on the description of digital resources and did not cater sufficiently for museums’ need to describe physical objects. CIMI also expressed doubts as to the ability of Dublin Core, a description standard, deeply embedded in the bibliographic tradition, to adapt to the specific needs of the museum community.

CIMI claimed the museum community had a need to provide contextual information, about people, places and events and the relationships between them and other objects. At the time, CIMI encouraged the use of the DCMES for basic inter-operability, but “in conjunction with a domain specific model that can be used by museums sharing rich, complex records” (2).

In this paper, we argue that much of what CIMI reported is still valid, although Dublin Core has undergone significant changes in the last five years and many of the short-comings reported in 2000 have since been overcome. There is much more richness available in the DCMES as it is now defined and used in practice. In particular, CIMI’s experiment relied heavily on the dumb-down principle for inter-operability. This rule has been degraded and is largely replaced by the technique of cross-walks and the development of application profiles. In addition, DC metadata can now be represented according to the Resource Description Framework (RDF) (3).

In the Quinkan Matchbox Project, a recent Australian experiment, qualified Dublin Core metadata was tested in a new museum-creation context. In this evaluation, techniques associated with using an application profile (4), a facility for cross-walking between metadata sets, RDF and recent developments in practice, were all assumed.
2. The Quinkan Matchbox Project

The Quinkan Matchbox Project has a content management system to support a small Indigenous community of Tropical North Queensland (Australia) in the management and preservation of its cultural knowledge (4). It is designed to contribute to the ‘virtual’ repatriation to the Quinkan community of cultural materials and intellectual property that are dispersed in a vast array of national and international organisations. The museum focuses on a culture that relates to at least 100,000 cave paintings and is of more than physical artefacts. It is a virtual museum of Quinkan culture.

The Quinkan Matchbox Project is inspired by the Digital Collective model (5) which encourages organisations and individuals alike to enter into partnerships with communities, and contribute physical and digital materials and their descriptions to a locally managed system.

A central metadata application profile (MAP) for the Quinkan Matchbox assists in the local description and classification of distributed resources. The Quinkan profile supports the import, integration, and export, of metadata records produced by a variety of intellectual communities for a variety of purposes with tolerably low levels of loss. The Quinkan MAP is Dublin Core compliant and interoperates with records of institutions (or classes of institutions) most likely to enter into partnership with the Quinkan community. In general, these are other museums.

The Quinkan MAP also reflects the Quinkan community’s view of cultural heritage, as best it can be understood, with local usage elements, element refinements and vocabularies. It uses qualified Dublin Core and its architectural features as well as recommended practices.

3.1. Quinkan country and culture

‘Quinkan Country’ is located in Tropical North Queensland, in the southeast region of Cape York at the north-eastern corner of Australia. Its centre is the small township of Laura with 60 or so residents (mostly Indigenous). The estimated 100,000 ancient paintings and engravings are scattered in natural galleries and sites throughout the lush savanna and high escarpments of Quinkan country. European settlement and the gold rush of the 1870s caused enormous disruption to traditional life. Stories are told by local Elders that bear witness to the violence, dispossession and displacement endured in the first 200 years of European settlement. Many are corroborated by European -style historical records. Some of these events are also represented in Rock Art (local name for cave paintings).

The few local Indigenous families represent different language groups and include descendants of the original inhabitants. Despite the past tragedies, the community’s sense of connection with ancestral places is strong. The Quinkan Matchbox Project was established in partnership with this community.

3.2. A real and a virtual museum

Quinkan cultural ‘property’ is not limited to Rock Art. A lot of physical and intellectual property is held in a variety of forms in remote cultural and government institutions, such as museums, libraries and archives scattered around the world. Intangible cultural property is embedded in local knowledge and human experience and in the minds of former community members who have left the area.

The Quinkan Project embraces the notion that repatriation and preservation of cultural content (artefacts, cultural and intellectual property) are vital to strengthening or revitalising local Indigenous communities (6). Online systems offer an alternative to physical repatriation for those communities that lack the means to support and maintain a hard-built cultural infrastructure. Although there is now a small interpretation centre in Laura it is not resourced to manage a collection of physical objects. In any case, the paintings are scattered over hundreds of square miles and the other objects are distributed around the world.

3. Resource descriptions

Current Dublin Core practice and principles use the extensible architecture of Dublin Core. It allows communities to develop their own element sets of qualified DC and additional elements, all selected from well-formed metadata schema, and work with local element sets called ‘metadata application profiles’ (7).

There are many ways to select a schema, or elements from one, for a specific application profile. Digital libraries often select the schema that best reflects the nature of their resources and their cataloguing goals. An application profile may result from the combination of ‘cross-walks’ from one metadata standard to several others to describe resources in rich ways and make them available through a variety of portals and search interfaces (8). Other communities may proceed in reverse, developing discipline-specific standards and then extracting a subset of their data to map to DC elements (9). In other cases, designers first consider the kind of functionality they want to offer users and then tailor their profile accordingly (10). Often where there are established description standards, communities can
translate them into application profiles containing Dublin Core elements.

Gill claims the MARC family of standards has supported the library community for the description and control of bibliographic materials. Cultural repositories, by comparison, do not share a common descriptive standard. In the cultural domain, there is, in fact, a plethora of descriptive models “that differ by institution type, collection type, curatorial methodology, level of detail and granularity, and intended applications and audiences” (11).

3.1. The Quinkan MAP

In the early stages of development of the Quinkan MAP, there was a review of standards and metadata in use in relevant cultural heritage and library domains, both in Australia and internationally. The work focused on institutions with collections of the kind of objects (real and digital) likely to be of interest in a Quinkan virtual museum.

The Quinkan MAP development followed the ‘ARH’ process: aggregation, rationalisation and harmonisation (12).

The Institute of Aboriginal and Torres Straits Islanders Studies’ (AIATSIS) (13) catalogue is an especially rich source of bibliographic records about Quinkan materials, including press clippings, sound recordings, collections of slides and bequests of archaeological charts and drawings. AIATSIS has a large user base of Indigenous researchers and extensive experience of their interaction with the catalogue and has adjusted its MARC-based bibliographic records accordingly.

The Environmental Protection Agency (EPA) (Queensland) uses the Indigenous Site Card Database (14) to record information about material heritage in a standardised form. The Register of the National Estate (RNE) describes in excess of 13,000 places of natural, historic and Indigenous significance (15). Such records are rich with data about the Quinkan culture that museums would typically want to collect. This data is specialised according to the discipline of the research reported. There is no simple way to merge it without significant and intolerable loss of information that would be of interest in cultural interpretation. The data is not always compatible: there is no established taxonomy of such data because, as is typical of museum information, there is not a single unifying perspective. As discussed below, the richness of the metadata from the various disciplinary and institutional perspectives is the major asset of a virtual museum, so rather than deplete it, the Quinkan MAP, like other museum systems, should collect it.

In the third step, the harmonisation stage, the Quinkan MAP v. 5.0. took advantage of several recommended cross-walks from and to DC (16-18).

4. Specific requirements of cultural resources

This section is based on a review of cultural resource descriptions proposed by a number of museum and heritage communities. It includes some of the specific requirements and problems related to describing resources in the Quinkan context.

4.1. Role information

Library and museum communities are often interested in recording the role played by a person in the creation of cultural materials (such as author, illustrator, sculptor, and engraver). Typically, archaeological charts and drawings are recorded by project teams, sites are surveyed by a variety of people ranging from local resident to park ranger, and objects without an identified creator may have a collector and a donor whose roles and identities represent essential contextual and, often, identifying information.

DCMI does not have recommended qualifiers for the Creator and Contributor elements but recommends the use of application profiles for this purpose. At the time of writing, the method for using role values as element refinements is awaiting consideration by the DCMI Usage Board after recommendations from the DCMI Agents Working Group (19). The problem for the Usage Board is that, while roles may be interesting, they are descriptive of the person and not of the resource. On the other hand, libraries and museums consider them to be relevant to the description of the resource.

The new DC Libraries Application Profile supports the use of role values as semantic refinements for Creator, Contributor and possibly Publisher (20). When drafting the Quinkan application profile, the need for recording the role of ‘agents’ who played a role in the life-cycle of the resource described was often an issue but most values suggested by the relevant catalogues and datasets scrutinised in relation to the Quinkan MAP found a match in the MARC List of Relators and Roles (21). The Library Application profile has established a model that could be used for museums.

4.2. Biographical information

Recording biographical information about artists and craftsmen is another special requirement of documenting resources in the cultural heritage domain. The boundary between empirical and interpretive statements about resources is tenuous. The CDWA has a set in which extrinsic or contextual information
about resources is treated as intrinsic information (22). The J. Paul Getty Trust has developed the Union List of Artists Names (23) as an authority file for the Creator category. It contains more than 255,000 artists’ names, complete with biographical and bibliographic information. Authorities are maintained separately from records, but are linked to them. This means they only need to be created and maintained once within the system. There is no mechanism to do this in the DCMES.

As with roles, the issue is about structuring the data to be recorded and what is being described but in the case of biographical information, it is even less about the resource and more about the person.

In 2004 the DCMI Agents Working Group was re-chartered to develop a core set of elements for unambiguously describing agents (people or groups) associated with resources. The work plan included the development of an identifier scheme to identify unambiguously a specific individual agent (24). The draft proposal for people emphasises the description of contemporary agents in a business-like setting. In the context of museum-relevant agents, especially for Indigenous cultural resources, it is not always possible to identify agents in such a way. They may be known only by a nickname or by a specific style. The emphasis on the agent’s description in this context is less on contact details and more on related dates and places, affiliation to language, ethnic or cultural groups, life roles and other biographical details. The CDWA set suggests an elaborate structure for the description of people, including place and date of birth and death, dates of activity, life roles, nationality (or ethnicity) and gender.

The Agents Working Group approach will need to be broadened if it is to cater adequately for museum resources. Again, as rich metadata is data for museums, while maintaining interoperability, they need to preserve as much of the data as possible.

4.3. Type

The variety of types of objects of interest to museums is wider than for typical libraries. Such values as “work” and “image” are included in the Visual Resources Association Data Standards Committee (VRA) Core Categories 3.0 (25). In the context of the VRA Core 3.0, a work is a physical entity that exists, has existed at some time in the past, or could exist in the future. An image is a visual representation of a work. A similar recommendation was made by CIMI with the suggested values “original” and “surrogate” (26), as it was felt that the DCMI Type vocabulary failed to address the need of the museum community because it classifies people, places and organisations alike as ‘physical objects’.

CIMI also proposed the addition to the type list of “party” and “place” to improve resource discovery of people and sites. CIMI found member museums often used ‘natural’ or ‘cultural’.

The future resources of the Quinkan Matchbox are likely to include records about archaeological sites and other locations of significance to the community, and also visual documentation of these locations. For example, the site cards produced by the EPA focus on the resources and describe their location, dimensions, orientation, spatial coordinates and more. For each such site, it is anticipated that related visual documentation will be of many types and in many formats. The draft Quinkan MAP includes extra values for the Type element with a view to assisting in the differentiation between records about ‘real’ places or objects and records about images of these places.

Many Indigenous authors have tried to explain ‘country’ as the location of Indigenous culture. They stress the dichotomy between natural and cultural does not exist for Indigenous people as it does in the Western worldview. For this reason, the value-pair natural or cultural is not in the draft Quinkan MAP.

The original perceived shortcomings of the DCMES can now be adjusted easily with application profiles, as was done in the Quinkan MAP.

4.4. Localisation of metadata

Indigenous knowledge is embedded in communities and is unique to a given culture, location or society (27). Tradition is rarely frozen and its vitality depends largely on its ability to evolve and adjust to changing contexts (28). To be useful, metadata must be suitably customised or ‘localised’ to reflect the current local worldview. It must also be ‘flexible’ enough to discourage ‘enduring representations’ for public consumption, or constraining cultural interpretation. Ideally, metadata should follow the evolution of culture in real life. This is particularly true of metadata used in a digital collective model and where metadata is used as data, as in the case proposed by the Quinkan Matchbox Project.

The Quinkan MAP is likely to remain a work-in-progress, enriched by each round of consultation and each new partnership, for many years. At this stage of its development, the MAP is established as a base from which to define more local rules and imagine more facets to represent the Quinkan information.

The early, simple DC element set could offer localisation and locally-specific details while maintaining interoperability by using the dumb-down model, in which such details were usually lost. Now, qualified Dublin Core provides for details and local specificities and these can be maintained, particularly by representation in suitable Resource Description Map (RDM).
4.5. Date and Coverage

Matters of spatial and temporal coverage present a specific challenge in the museum world. Dating cultural materials precisely is rarely possible and refinements indicating the degree of accuracy of the date values must be considered, in keeping with the recommendations of the Core Data Standard and the CDWA: date absolute, date earliest, date latest, or circa. Many museums resort to using terms, preferably selected from a controlled vocabulary, to express the culture or style of a work of art (“Italian Renaissance”, “Art Nouveau” or “Aztec”). Early use of the DCMES focused on interoperability being achieved by the use of limited vocabularies and formats and data-rich descriptions of interest to museums were not encouraged.

In the Quinkan Matchbox context, Australian archaeological periods are named differently from their European counterparts and have their own distinct chronology. Data entry guidelines are required to assist users of the Quinkan Matchbox in formatting relative dates according to the chosen standards, so that they remain machine-readable.

Matters of date and spatial and temporal coverage of a resource also will need special vernacular to reflect the local usage. The literature stresses that Indigenous people are not necessarily as interested in chronology and linearity of time as archaeologists might be. Barker and Gaston (29) have described how people in the Whitsunday Islands (Queensland, Australia) use broad categories for the past (“before people”, “long time before”). Baker (30) says the Yanyuwa people (700 km east of Darwin, Northern Territory) classify their history in a succession of ‘times’: Macassan times, wild times, police times and more. These temporal divisions were useful to organise the “Land is life” Web site of the Yanyuwa people and culture online (31). The Quinkan Project anticipates similar themes emerging from work with the community.

The Yanyuwa people, like many Indigenous people of the Tropical north of Australia, divide the year into five, seasons (30; 32). This has also been documented in the Quinkan region (33). It can be mapped to the months of the ‘European’ calendar. This short scheme has been included in the Quinkan MAP to make both the discovery and classification of resources more intuitive to local Indigenous people, but retains a discreet correspondence to a twelve-month cycle.

There is now no difficulty in associating the necessary date details with the DC metadata elements, as done in the Quinkan MAP.

4.6. Language

DCMI recommends specifying the language of a resource using RFC 3066 (3), with ISO 639 (34). This ensures interoperability with other systems using the same taxonomy.

Unfortunately the ISO 639-2 list of languages recommended by DCMES does not contain any Australian Indigenous languages (34). This suggests it might also not contain many other Indigenous languages.

In the Australian Institute Aboriginal and Torres Straits Islander Studies’ catalogue, the preferred term for a given language name is linked to secondary records containing spelling variations and naming variations (e.g., 60 variations for one particular example). This maximizes discoverability. AIATSIS’ choice relies primarily on the work of anthropologists and linguists such as Tindale and Dixon, who have built major linguistic typologies. The files relevant to the languages and cultural groups of the Quinkan region and its surroundings are not available to the Quinkan Matchbox. This means searches based on languages and cultural groups will need special attention. It is likely that special searching aids, like maps or lists, will be needed for users.

DCMES allowed early for the identification of the language of metadata content but many of the problems associated with multi-linguality will not be solved in the very near future.

4.7. Subject and Keywords

Finally, the DCMES allows a significant amount of flexibility in the use of subject and keywords. Prior to the development of qualified Dublin Core, there was a facility for naming a scheme such as the Library of Congress Subject Headings (LCSH) or ‘keywords’ that were entered as free text.

Global subject indexing tools like the LCSH were developed for particular communities and reflect their values. These tools are now being considered for use in many other contexts but they have received criticism for being imbued in Western concepts and prejudice and undermining minority ‘worldviews’.

In Australia, LCSH and other similar indexing vocabularies have been accused of assisting in the creation of an unwarranted “otherness” in their classification of Indigenous resources. Moorcroft has noted that they often dilute the real nature of the content of certain resources under polite and comfortable headings (35; 36). The choice of a more ‘localised’ indexing vocabulary may be more appropriate and increase the discoverability of the Quinkan Matchbox resources by providing a greater number of entry points and also by locating them...
within the Australian and Indigenous context. The Mura Gadi service produced by the National Library of Australia uses a (modified) version of the Australian and Torres Straits Islander Thesaurus (37) to re-index Indigenous resources held in its collections. Terms like “Land rights”, “Terra Nullius” and “Tent embassy” are uniquely Australian.

The Australian and Torres Straits Islander Thesaurus (38), created to redress the imbalance and oversight of the commonly used controlled LCSH vocabulary, is currently the preferred source of subject terms for the Quinkan Matchbox. It is envisaged that it will be refined with the addition of local and regional terms, suggested by users or extracted from the resources themselves.

5. Dublin Core 2005

In general, the difference between DCMES in CIMI times and today is both significant and favorable enough to enable its use now for rich descriptions of museum collections. Qualified Dublin Core, application profiles, common use of XML and RDF, have all expanded the potential of the DCMES. Recently the adoption of an abstract model has clarified the DCMES architecture and so helped clarify its use.

5.1. Metadata as data

In the case of museums, metadata is not just for discovery although that is an important use for it. In terms of discovery, many argue for a simple ‘who, what where, when and how’ taxonomy for museum discovery. Other Indigenous culture projects tend to shy away from classical cataloguing and let the culture speak out from a Website they create with Indigenous input. The Quinkan catalogue follows the approach of mainstream museums: the original CIMI dumb-down principle is now avoided by the aggregation of metadata into facets, possible the “when, who, where” set, or others, according to the immediate purpose for which the metadata is being used. The provision of multiple entry points into the metadata, for example, means that it can be re-used in a variety of combinations ranging from a simple set of facets for discovery to a complex set for researchers.

The DCMES provides a framework for the rich data that often accompanies museum objects, such as a rock art painting in the Quinkan case, while making it easy for a cataloguer to add metadata to a document that is more bibliographic and so does not have many of the characteristics that might be of interest in the case of the painting. Slim and rich records can co-exist. Museums seek rich data about their collections while others aim for a minimum amount of metadata only for discovery.

The combination of information sources, a major goal for museum work, and the ability to extend this to automated harvesting of information from a range of external sources, means that much of the information that will end up in the metadata is data at its origin. For instance, the EPA has detailed records of painting sites: to enter such information as metadata would be extremely tedious but to be able to harvest it from an existing database, where it has been collected as the primary work of a ranger, is easy. Also, while it would be possible to record the existence of the EPA’s database and describe it as a collection, it is the integration of that information into a wider pool of information that makes it most useful in practice in the new museum.

As metadata records are such a significant part of the data of the virtual museum, their flexibility is exploited further when they are in a form that supports the many voices that now are often sought to interpret cultural heritage.

Annotations of metadata records are becoming common in the museum world. One incentive for using annotations is that multiple languages are often relevant to collections. It is also the case when the collection relates to a culture with no written tradition and cultural experts are asked to talk about their culture (to annotate records ‘verbally’, for example).

6. Further developments of interest to museums

Today, the use of subject terms is no longer just a choice between free text, controlled vocabularies, or thesauri. Google has set a new standard in the use of full text for descriptions of resources by developing many algorithms that seemingly satisfy most resource discovery requests. Where Google’s approach is inadequate, metadata is used to make available human judgments about resources that often fill the gap to solve the discovery problem.

The recent significant development in technology that provides living (dynamic) ontologies extends the value of rich metadata. Ontologies are vocabulary systems that grow according to rules, so they evolve like living languages. In addition, the growth of the Semantic Web is increasing the vast range and style of museum vocabularies that can be accommodated.

Increasingly, museums enrich basic empirical statements about resources (name, date, size, provenance) with statements of significance. While empirical statements were well served with controlled vocabularies and standard notations, adding richer descriptions means that multiple, more subjective, interpretive narratives can be recorded and related to one another and to the resource, using an appropriate
syntax. This is of particular value for a catalogue of Indigenous culture.

In their introduction to the ATSI Thesaurus, Moorcroft and Garwood note that Aboriginal languages focus on the qualities of things – on people, and how they relate to each other, on land and spiritual ideas, and the connections between them. The interconnected nature of people, places, objects and concepts in such context are not likely to be described effectively using rigid, uni-dimensional organisational principles (39).

It is not so much a choice of which hierarchical taxonomy or thesaurus as it is a choice of style – taxonomy, facets or rich descriptions that support the multiplicity of points of view from which a resource can be described, questioned and interpreted.

7. Conclusion

With its focus on resource discovery, Dublin Core does not yet cover the vast range of functions metadata can enable. Dublin Core has evolved and matured since CIMI first demonstrated some of the shortcomings of DC metadata when applied to museum resources. The Quinkan MAP takes advantage of the extensibility principle and extensible architecture of Dublin Core to include both resource-specific and localised extensions. It is designed to interoperate without loss (or at least with only a tolerable level of loss), with other profiles used by organisations most likely to enter into partnership with the Quinkan community.

The additional practices now recommended for DC users allow for its use in the museum context. Particular demands borne out of the Quinkan project have shown, however, that Dublin Core itself still has limitations when it comes to describing people, places and events with sufficient precision.

References


from http://dublincore.org/groups/agents/
http://www.loc.gov/marc.sourceforge/relator/relatorlist.html
the Description of Works of Art (CDWA), from http://www.getty.edu/research/conducting_research/standards/cdwa/
http://www.getty.edu/research/conducting_research/vocabularies/ulan/
requirements for describing agents (Working Draft 2), 2004, from
http://dublincore.org/groups/agents/agentFRdraft2-2.html
http://www.vraweb.org/vracore3.htm
and Indigenous knowledge: towards an
Indigenous knowledge information system. Paper
presented at the Information Technologies in
Educational Innovation for Development:
Interfacing Global and Indigenous Knowledge
(Sixth UNESCO-ACEID International
Conference on Education, 12 – 15 December
2000), Bangkok, Thailand.
28. Merlan, F. (2000). Representing the rainbow:
Aboriginal culture in an interconnected world.
knowledge and the tyranny of text: archaeology,
ethnohistory and oral traditions in Australian
archaeological interpretation, 2003, from
geography of Australian contact history. In E.
Stratford (Ed.), Australian Cultural Geographies
(pp. 25-47). South Melbourne, Vic.: Oxford
University Press.
our art, our Quinkans. Laura: Ang-gnarra
Aboriginal Corporation.
34. (2002). ISO 639-2 - Codes for the representation
of names of languages. Alpha-3 code (ISO 639-
2:1998), 2004, from
headings. The Australian Library Journal, 41(1),
40-45.
Aboriginal and Torres Straits Islander Thesaurus.
38. (2002). Mura Gadi: a guide to manuscript,
pictorial and oral history materials relating to
Aboriginal and Torres Strait Islanders in the
National Library of Australia, 2004, from
Matchbox Project: challenges in developing a
metadata application profile (MAP) for an
indigenous culture, 2004, from