Do We Need Application Profiles?
Reflections and Suggestions from Work in DCMI and ISO/IEC

Liddy Nevile
(Retired)
Australia
liddy@sunriseresearch.org

Eva Méndez
LIS Departement / iSchool
Universidad Carlos III de
Madrid, Spain
emendez@bib.uc3m.es

Abstract

In this paper, the authors question the role and naming of ‘application profiles’ (APs). It is not a research paper but aims to foster a discussion that the authors think is pertinent. Both have been involved in the development and use of application profiles for some considerable time. This paper does not provide answers but aims to raise issues for others’ consideration. Essentially, the issues show that communities can share work easily through the interchange of APs but suggests that greater precision in their naming would be useful, and they may not always be necessary given the current state of RDF technologies.

Keywords: application profiles; metadata; metadata schemas; APs; MAP; RDF; discussion

1. Introduction

When someone you really respect makes a comment, even casually, it can fester for days. How about, “Why do you want an application profile? They are not necessary…” This comment was made in the context of developing a standard for Sub-Committee 36 of the Joint Technical Committee 1 of the ISO/IEC, a committee working on standards for ‘IT for Learning, Education and Training’ (ITLET). The target standard that had already been adopted and even mandated in Europe (ISO/IEC N24751) concerns accessibility of resources but was being significantly revised. The context included the development of a new comprehensive Metadata for Learning Resources (MLR) standard for ITLET (ISO/IEC N19788). The latter standard is very strictly Resource Description Framework (RDF) compliant and it is for education, so it also offers support for Learning Object Metadata (LOM) users, and for many in other DCMI related communities.

A similar comment was made by a student at the end of a course on “Metadata and Vocabularies”, after reading all the course material, recommended bibliography and so forth. He asked, “It is so common as it seems, the creation of so many application profiles? It seems that every single project of digital information service requires its own “customized” metadata schema? It has to be like that? It is not against the standardization that you said surrounds the metadata? On top, currently there are several schemas applicable to different projects so, I am wondering if it would not be enough choosing one of those standardized schemas.”

So, in different contexts, both authors have heard the same Why do you want an ‘application profile’? or Do you really need yet another ‘application profile’? The comment seems worth consideration in the context of another Dublin Core Metadata Initiative (DCMI) conference and set of tutorials, including one on application profiles (APs). In addition, the profiling process has expanded, given the proliferation and availability of RDF and Semantic Web technologies.

As Murtha Baca from Getty said, metadata standards are sometimes like toothbrushes, everybody thinks that they are a very good idea, but everybody prefers to use their own (Méndez, 2007). In this paper, the authors consider the role and naming of sets of terms for description of entities, of ‘application profiles’.
2. Metadata Application Profiles (APs)

A Metadata Application Profile, Metadata AP, or just MAP or AP can be understood from a number of definitions ranging from the more general one in Wikipedia to the more specific ones in the DCMI context. Wikipedia1 defines an AP in the domain of ‘computer science’:

an application profile consists of a set of metadata elements, policies, and guidelines defined for a particular application. The elements may come from one or more element sets, thus allowing a given application to meet its functional requirements by using metadata from several element sets - including locally defined sets. For example, a given application might choose a subset of the Dublin Core that meets its needs, or may include elements from the Dublin Core, another element set, and several locally defined elements, all combined in a single schema. An application profile is not complete without documentation that defines the policies and best practices appropriate to the application (Wikipedia, 2015).

At the time of writing, the Guidelines for the Dublin Core AP say:

A DCAP includes guidance for metadata creators and clear specifications for metadata developers. By articulating what is intended and can be expected from data, application profiles promote the sharing and linking of data within and between communities. The resulting metadata will integrate with a semantic web of linked data. To achieve this it is recommended that application profiles be developed by a team with specialized knowledge of the resources that need to be described, the metadata to be used in the description of those resources, as well as an understanding of the Semantic Web and the linked data environment (Coyle and Baker, 2009).

Thomas Baker et al. (2001) noted: “It is rare that requirements of a particular project or site can all be met by any one standard ‘straight from the box’. Different metadata implementations may have different perspectives. Different information contexts, different content or different user requirements can motivate the creation of a Metadata Application Profile for local purposes. MAPs are the performance of the “Think global, act local” principle applied to metadata in domain-oriented digital information services. In fact, an early principle that drove the development of the DC Terms was that communities were likely to have domain relevant needs that may have little value beyond their context. On the other hand, they were surely interested in sharing their descriptions and therefore their term sets (APs) because that would assist with interoperability. This driving principle was embodied in the slogan ‘global interoperability and local specificity’ and cited many times in the early days of DCMI. In fact, APs were nurtured in a context where it was well-known that not every metadata system would be the same.

It should be pointed out that the work of Hunter and Lagoze (2001) led to the OAI (Open Archives Initiative) developments that partially solved the problem of sharing relevant ‘global’ metadata while the APs were conceived to solve the localisation problem.

In practice, it seems that nowadays communities develop APs for their domain of activity but it is safely assumed that when these are implemented locally, system developers choose what is of use from those community APs and locally they will, for sure, add some terms for local use (such as collection acquisition dates, or benefactors, for example). This practical approach to the use of APs has been described as a process in which the APs are used as “metadata building blocks” (Zeng and Qin, 2008). An AP may also be based on single schema but tailored to different user communities (Chan and Zeng, 2006); examples include the DC Library Application Profile (DC-Lib) used by libraries and library-related projects and applications or LOM-ES that explains the use of the Learning Object Metadata elements by the Spanish speaking community.

1 The authors are quoting Wikipedia deliberately because that is where most people find meaning for such expressions. They are well aware of the many detailed, carefully defined definitions of application profiles that have been developed by authoritative entities, communities, in academic papers, etc.
Another way of dealing with this issue has been shown by the DCMI education and accessibility communities. Curiously, in concurrent meetings of the two groups some years ago at a DC Metadata Conference, the notion of metadata ‘modules’ was raised by those communities. They recognised that they had specialised needs but that what they wanted was just a small set of additional terms that could be used alongside the more general set of terms, or added to an existing AP.

So we start our conversation with local resource profiles, community resource profiles and modules for resource profiles in mind. Let us now analyze two particular AP contexts: DCMI, and ISO/IEC JTC1 ITLET.

3. DCMI context

3.1. Metadata for Education in the DCMI context

The aim of the first DCMI application profile, designed for education back in 1999, was to find a way the small DC element set could satisfy the needs of a specific community. Until then, DCMI work had been focused on developing a general set of elements for everyone to share. The value of the application profile, a slightly extended set of elements, was that it increased the value to a particular community by putting the focus on the properties of interest, following the DCMI idea of promoting ‘global interoperability and local specificity’. This work was undertaken in the development of the Victorian Education Channel in Australia (Nevile, 2008, p 126).

The element set extension was done with the assistance of the then Director of the DCMI who considered three factors important. Any new term should:
- not redefine terms,
- not duplicate terms, and
- follow the dumb-down rule. (Nevile, 2008, p. 127)

Significantly, the new terms were to further describe the attributes of a given resource.

The exercise helped broaden the use of DC elements. The idea of application profiles was formalised in a paper written and published shortly afterwards by Rachel Heery and Manjula Patel (2000) where they specifically attached the concept of AP to data elements from different namespace schemas being combined by the implementor in a way that was optimized for a particular local application. Heery and Patel explained that application profiles are useful as they allow the implementer to declare how they are using standard schemas (APs). Thus there was recognition of a community developed schema (AP) and a local AP, often built from a combination of components of other APs. Again, the main aim is clearly to maximise global interoperability and, at the same time, local specificity. It was and is still also to enable better descriptions of an entity for the process of matching user requirements to available resources (or services).

Unfortunately, it seems in hindsight, the name ‘application profile’ stuck, without clarity about why. A number of different term sets were given the same name. For the purposes of this paper, the authors have distinguished between sets of terms for describing resources based on:
- agreement among a wide community for publication, and
- relevance to a particular context.

These sets can then be further defined as being determined to cover.
- fixed attributes of particular available resources but also, incidentally,
- user specified attributes of resources commonly thought of as search criteria in the discovery process.

The second distinction helps clarify that resource descriptions are always potential search criteria, or what has been described in other contexts as user needs and preferences (Nevile, 2005a, 2005b). Put simply, a resource provider might describe the date of publication in a
standard way and so a user can search for a resource with that date of publication in a corresponding way. There is nothing new in this, but the focus for a long time seemed to be on resource description, usually agreed among resource providers or organisers, and the use of the profile as search criteria was simply covered by saying the main use of the metadata was discovery.

This distinction is also useful because the growth and ability of 'search engines' that do not depend on what is commonly understood as metadata, or rather the lack of visibility of such search criteria, has led many to believe that search engines don’t use metadata. Hopefully this myth has been rapidly and forcefully debunked recently by the spectacular growth and adoption of the work of schema.org. These new sets of metadata terms for ‘all-the-web’ retrieval systems is, at some point a ‘déjà vu’ for the authors, since schema.org revives the dream of qualified and precise search in the Web through metadata, like Altavista tried in the 90s. It is not the case even now that search engines necessarily use metadata in the same way, or the same metadata, as more formal traditional systems, but at least there is now an open dialogue between the two discovery system providers.

At the same time as the use of APs was evolving, the DCMI was working on what emerged as its 'abstract model' (DCAM) (Powell et al, 2007). Later in DCMI's life, Nilsson tried to find agreement between the DC metadata and LOM metadata in the educational context. He found that very different models led to very different forms of metadata and they could not be matched, so lossless interoperability was not possible. In general, he showed how difficult it is to match metadata from different structural models and argued for metadata to be interoperable it must be developed at least using compatible models, and developed a structured model to explain this (Nilsson et al, 2008). Following this work, metadata interoperability is considered by levels of interoperation. This model is illustrated in the following figure.

![Diagram](image)

**FIG. 1:** the “Singapore Framework” for interoperability (Nilssen, et al, 2014)

At this time, the domain for description was always just considered as a ‘resource’. (Perhaps significantly, DCMI did not distinguish between potential values that are now known as literals and non-literals.) But there was a rule, known as the one-to-one rule, which limited a description to a single resource. In the description or discovery of a resource, one might want also
to describe attributes of the person who made it, or the use to which it has been put. Curiously, the DCMI community overlooked that an image was a distinct entity, and for accessibility purposes may require a comprehensive description for those who could not see it. On the other hand, DCMI required a description of the resource and a second description of the person with the two descriptions linked by the term ‘relation’.

Clearly, DCMI was a pioneering agency and the work was breaking new ground. The structures chosen were the best known at the time.

The integration of an image, or other object, into a resource was easy to live with when resources were published in what we might be described as a single entity form, despite really being a compilation (even including redundant parts like a long description of a diagram, tagged in the HTML as a longdesc). But today many resources are compiled ‘on the fly’ according to a wide set of requirements based not only on subject matter but location of the user, a number of previously exhibited behaviours, and more.

3.2. Metadata for Accessibility in the DCMI context

A short time after the forming of a DC Education Community, a DC Accessibility Community was formed. In the case of accessibility, a potential user needs to know such things as if there is a text alternative for an image, if a service can be controlled using only a keyboard or an assistive technology driven by a keyboard, or speech, perhaps. This means that a resource might need to take redundant forms, may not have all components assembled in a fixed format, and might need to be accompanied by an associated but not incorporated description of itself.

For more than a decade, the use of metadata to help solve the problem of inaccessibility of resources for people with disabilities has been pursued. There are guidelines for making resources that are, supposedly, accessible to everyone following what is called ‘universal design principles’ (W3C/WAI Web Content Accessibility Guidelines known as WCAG). Unfortunately, these guidelines are rarely followed successfully and even if they are, they do not satisfy everyone’s needs simultaneously (e.g. Petrie & Bevan, 2009).

For some time, a term proposed for describing the accessibility of a resource was deemed unacceptable for technical reasons. Finally, a single term was adopted but the original hope that accessibility would become an important part of a DC metadata statement was not supported. This led to the work being taken to the ISO/IEC JTC1 context.

4. ISO/IEC context

4.1. Metadata for Education in the ISO/IEC ITLET context

ISO/IEC N19788 is the Metadata for Learning Resources (MLR) standard. The MLR is very detailed in its ways of defining application profiles (Part 1) and includes several APs. It has many parts and bridges earlier practices in both ISO/IEC’s provision of what we now call metadata terms, and the practices associated particularly with older database systems and the hierarchical structures of the LOM.

The interoperability of the MLR comes not just from working carefully with the earlier practices of describing electronic resources using a sort of document object model, but the fact that today not only resources as they have been traditionally known are to be described. There are people associated with the development and publication of resources; there are services and online communities. All of these things are connected in a web of digital descriptions so users can have very different points of contact with that web and it refers to objects that are digital but also physical or merely conceptual.

N19788 is not necessarily easy to read in its full form, but that is not necessary for its use. It is very helpful in that it does provide full explanations of its techniques. There has been considerable effort put into diagrammatic representations, examples in pseudo code, and bindings
that can be simply adopted and used. The interoperability of the MLR depends, in fact, on complexity that is buried in what appears as an elegant framework. 'Under the covers' techniques have been developed to ensure that terms are easily accessible online, that internationalisation is fundamental, and more.

The MLR offers global interoperability by specifying how to do a number of things but strongly supports local specificity in terms of extensions, options, etc. The MLR’s application profiles provide an initial set of core terms that can be used to describe educational resources, and these are effectively the 15 DC simple terms (limited currently by their domain to 'learning resources'). This application profile is in Part 2 of the standard. Part 4 of the standard offers a few specific terms to describe technical aspects of a resource and there is another application profile in Part 5 that has been developed by a community of educators to describe what might be considered the pedagogical aspects of learning resources.

The MLR goes on to include sets of terms for descriptions of, for example, the role of a person associated with the development of the learning resource, or of a resource that is, in fact, a metadata description of a learning resource. Such a term does not aim to support description of the resource itself, but an attribute of the person who has been described in association with a resource (or more particularly, the role of a person who has been so described). In this case, we think of the chaining of descriptions to link the various types of descriptions to form a web of information about the resource, significantly using RDF and data linking techniques, but it can be done however the user chooses.

The MLR has attempted to bridge the gap that emerged as technologies have moved from standard databases to more and more fluid systems. A considerable amount of what is in the MLR is concerned with this. The result, however, is that terms can be used and combined in very flexible ways.

4.2. Metadata for Accessibility in the ISO/IEC JTC1 Context

Today, many in the accessibility community have adopted the additional approach of profiling the needs and preferences of users, especially those with disabilities. The aim is for so compilations of resources to be matched to an individual user's stated needs and preferences so the resources are 'perceivable, operable, usable and robust' for them (WCAG). This, of course, means simply that if the relevant properties can be identified, they can be used for resource description by resource providers (or others) and for resource discovery (in search requests or automatically by systems). A delivered resource may not be accessible to another individual in the same form, or even to another user with a similar disability. The required form of the resource is for an individual user to define.

The AccessForAll approach, as it is known, was first proposed by Jutta Treviranus. It is simply a name for ways to describe an individual user's functional requirements for a resource that can be matched when a resource is being delivered. To metadata communities, it is a very normal metadata activity but somehow has not been recognised as such by a number of those who want it, and so, even after about 7 years of work, they have not managed to agree on what could be described as a simple AP - possibly all that is required!

A characteristic of AccessForAll metadata, as proposed, is that the attributes or properties of a resource are described using a set of terms which is the same set for a resource and for the search for the resource except that, in the latter case, there is no clear identity of the resource being described - its identity is being sought. That is, the same terms can be used but the identity of the resource is not specified in the latter set. It is appropriately described as a module of metadata. It challenges the idea that an AP is a set of terms that describes a resource. The description of the needs and preferences of a person, expressed as metadata, does not describe the person. In fact, a single person may have a number of stored accessibility modules describing the functional requirements that they use at different times, in different locations, and even according to different purposes.
Giving the terms for a search a different name from the terms for a description should settle this easily. It should not be a show stopper. It means simply that the identifier should be optional, whereas it has always been mandatory for a DCAP. There is nothing special about a module that describes attributes related to accessibility: the same idea can be used for many types of customisation of resources useful to anyone. This is generalisable as ‘inclusion’, the preferred way of avoiding discrimination.

The values set for the term set is, itself, an entity and that, of course, can have an identity in the form of a URI, or otherwise. It can have lots of other attributes as well and they too can be described in a value set.

The AccessForAll metadata approach has received significant funding for many years and is often considered to be exemplified by the project known as GPII (Global Public Inclusive Infrastructure). Sadly, despite the funding and academic papers and other peripheral successes, the simple matter of providing an ‘application profile’ for accessibility has not yet been achieved.

The AccessForAll idea is to have a profile of the needs and preferences of an individual user (could be anyone but should, at least, be inclusive of any person with disabilities) and to match those needs and preferences, strictly described as functional needs, to resources. If a resource is well-developed and has available components, possibly redundant, that can be combined to make all its essential content available to the individual user, the useful combination should be delivered.

If components are not accessible, for example an image is not also described in text, an alternative resource might be located or created to serve this purpose. The useful component can be linked to the original using the metadata. The whole matching exercise is known to provide what is an ‘accessibility service’ by constructing an ‘accessible resource’ and this can be a dynamic process, with cumulative accessibility.

5. RDF and APs

Developing the MLR (Metadata for Learning Resources) has provided an opportunity for re-thinking the original metadata and application profile ideas in a modern context, specifically in the context of an RDF world. Gilles Gauthier has provided an image of a web of descriptions as it might be for a particular learning resource of interest (Figure 1).
Using RDF triples, one could construct an impossibly huge web of descriptive triples for any resource, almost, but this is not always totally practical. If a resource description is to be useful, it may not matter from where the triples come, assuming they are reliable, and they can be all joined up but limited by a set of delimiters. These would be rules to say just how much information is wanted. The original map shown above is shown after a set of delimiters have formed the map that the particular user wants to work with (Figure 2).

So here is a question: Is what we see in Figure 2 an application profile? Is it possible for an application profile to be a set of delimiters? The MLR has lifted some of the restrictions earlier encountered in a way that is being done by many others. The current authors would like to suggest that the focus on 'application profiles' that has been useful in the past may benefit from some sort of re-thinking in the light of such new possibilities.
FIG. 3: the RDF graph from Figure 1 with a set of delimiters applied showing how it is thus ‘contained’.

6. Discussion

During the last decade, there has been extensive work on the practices for using APs, perhaps well-exemplified in the work of Curado and Baptista (2013) that provides a carefully researched process for developing an AP. This current paper, recognises this work, best practices for system developers, but also the work of communities that have resulted in APs such as the LOM (Learning Object Metadata), the LRMI (Learning Resource Metadata Initiative), the MLR in the educational domain and more in other information domains.

The authors are not sure why APs have a special name, which is cryptic for people who are not metadata-literate. Isn’t an AP simply the set of terms that a user, whatever their role, chooses to use? Can’t the set of terms be just that, and the ‘development of an AP’ be simply recognised as the ‘use of metadata terms’, a general activity? Couldn’t the DCMI workshops on DCAP be known simply as workshops about how to use metadata?

Perhaps the value of the focus on ‘application profiles’ is that it helps people distinguish between the set of terms that can be used to describe a resource and the set of values for those terms. This difference is a significant problem for some people. For others, the idea of loose terms is a problem. They want to think of metadata terms as they thought of fields in a data base. They want them neat and controlled, probably in the same place, verified not just technically but by some authority. This is not the world in which metadata lives today. schema.org developers, some of the major search engine companies, have publicly stated that what they use will emerge according to what others use. Terms defined as part of schema.org that don’t gain popularity will be ignored. schema.org can safely adopt the position of allowing terms to see if they work. The unused terms will not do harm. Terms published and not used will be just that.

The authors’ interest in the name ‘application profile’ is perhaps also motivated by the fact that the expression is curious, even funny when translated into some other languages, and sometimes confusing.

The original specification of the domain of a DC term was very loose, simply ‘resource’. Given the different aspects of the resources we use today, there are a number of different parts of a resource that may need description so the domains will not be the same for all the descriptors in a useful metadata set. Already the capacity to handle this was proving a strength of Resource Description Format (now Resource Description Framework, RDF) metadata. That a resource may, in fact, be delivered in different forms or manifestations, according to user preferences, device types, etc., was not yet an issue.

The early RDF work anticipated that chaining of descriptions would be useful but, at the time, it was not well supported by software or implementations. RDF was not universally trusted in the late 90s - it seemed to be a folly for a small number of ‘academic’ types, semantic geeks and perhaps, as was often said, ‘people with comfortable shoes’. There was strong concern that it would go away so should be treated with caution. Time has shown something else. RDF is very
well established now and substantiates most of the Linked Open Data projects, but it is still relatively poorly understood. Many with a background in database work have a strong sense of structure and formalities that are not always compatible with RDF use. It is very hard for many to let go of those structures and leave organisation to the implementing systems. Nowhere has this been more problematic than in the ISO/IEC SC36 metadata work. There are representatives of technologically advanced nations who themselves do not endorse RDF and the Semantic Web. There perhaps even more who do not even know what is the difference between old databases and the Semantic Web... But today there are billions of RDF triples in use; hence the question being asked in this paper. Do we still need ‘application profiles’? Do we still need to use that name? Might we want to advise people just to ‘use metadata’ and even share it, or develop it collaboratively? Alternatively, might we want to be more specific and recognise the various kinds of metadata profiles? What level of standardisation or at least community endorsement does an AP need? Most of the metadata models are developed as ‘standards’.

Determining how to describe Japanese manga has offered a number of examples of resources that fall outside the norm. Manga, originally Japanese comic format (but often created with significant adult themes), is very different from standard literature. Like other comic series, characters re-appear in subsequent comics in a series. But more importantly, like literature, manga has a grammar. There are quite formal ways of signifying emotions and actions in manga (manga creators study for several years at the Japanese University of Manga in Kyoto). A useful way to think of manga is to compare it to ballet and other forms of dance. Metadata for the description of manga is complex - an elegant set of descriptive terms that includes the various attributes in the Functional Requirements for Bibliographic Records (FRBR) is the solution emerging from the work of Shigeo Sugimoto and his students in Tsukuba, Japan (Mihara, Nagamori, & Sugimoto, 2012). Such a complex AP would be beyond the average user to develop but once it has been established, it can be used easily. But how should such a set of terms be described? as an AP? What about calling it a ‘manga profile set’? Wouldn’t such a name be helpful?

Similarly, developing the MLR has been a very technically challenging exercise but the result is something that can be used by people with few computer development skills but good cataloguing skills, or maybe without them. The idea then is that the expertise to determine an appropriate set of terms and potential values for description of the wide range of resources may well be beyond the average user, but useful to them. But what is the MLR? it offers a number of APs but other terms as well. Metadata that mixes terms from the MLR and with others that conform to the MLR will be considered MLR metadata. Does it not make sense to talk about a set of educational metadata terms? In this case, given terms are defined both by text definitions (traditional term definition) and by constraints on RDF triples (newer term definition).

7. Conclusion

So, what is an application profile is not clear, according to the authors of this paper. It is a wide-ranging concept that is perhaps not even useful any more. Without reaching a conclusion, the authors hope to have stimulated some useful thinking and that some of the questions asked in this paper will lead to timely and useful discussion.

References


**Web References**

- RDF [http://www.w3.org/RDF/](http://www.w3.org/RDF/)
- schema.org [http://schema.org](http://schema.org)
- WCAG [http://www.w3.org/TR/WCAG20/](http://www.w3.org/TR/WCAG20/)