

SCHEMAS

Contract: N° IST-1999-10100

Forum for Metadata Schema Implementers

D63: Application Profiles for interoperability – a simple model for declaring “uses” of metadata terms

Document number:

SCHEMAS-GMD-WP6-D63-Final-20020207

General Information

Title	Application Profiles for interoperability – a simple model for declaring “uses” of metadata terms
Creator	Thomas Baker
Subject-Keywords	Deliverable D63; Application Profiles
Description	This paper proposes a simple model for declaring the metadata model of an application, with annotations on local usage practice
Publisher	Fraunhofer-Gesellschaft
Date	2002-02-07
Type	Text Manuscript
Format	application/msword
Identifier-URL	
Identifier-Document Number	SCHEMAS-GMD-WP6-D63-Final-20020207
Language	English
Rights	European Commission

Dublin Core Metadata in HTML

```
<META NAME="DC.Title" CONTENT="Application Profiles for interoperability – a simple model
for declaring ‘uses’ of metadata terms">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">

<META NAME="DC.Creator" CONTENT="Thomas Baker">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">

<META NAME="DC.Subject" CONTENT="Deliverable D63">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="WP6">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Subject" CONTENT="Application Profiles">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">

<META NAME="DC.Description" CONTENT="This paper describes an Application Profile model.">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#description">

<META NAME="DC.Publisher" CONTENT="GMD">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#publisher">

<META NAME="DC.Date" CONTENT="(SCHEME=ISO8601) 2002-02-07">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">

<META NAME="DC.Type" CONTENT="Text.Manuscript">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#type">

<META NAME="DC.Format" CONTENT="(SCHEME=IMT) application/msword">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#format">

<META NAME="DC.Language" CONTENT="English">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">

<META NAME="DC.Rights" CONTENT="European Commission">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#rights">
```

Dublin Core Metadata in RDF

```
<?xml version="1.0"?>
<rdf:RDF xml:lang="en"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:smes="http://www.schemas-forum.org/registry/schemas/SCHEMAS/1.0/smes#"
>

<rdf:RDF rdf:resource = "Application Profiles for interoperability">
<dc:title> Application Profiles for interoperability </dc:title>
<dc:creator> Thomas Baker </dc:creator>
<dc:subject> Deliverable D63 </dc:subject>
<dc:subject> Interoperability Profiles</dc:subject>
<dc:subject> WP6 </dc:subject>
<dc:description> This paper proposes guidelines for Application Profiles that declare how an
application uses metadata terms. </dc:description>
<dc:publisher> GMD </dc:publisher>
<dc:date> 2002-02-07 </dc:date>
<dc:type> Text.Manuscript </dc:type>
<dc:format> application/msword </dc:format>
<dc:identifier> SCHEMAS-GMD-WP6-D63-Final-20020207</dc:identifier>
<dc:language> English </dc:language>
<dc:rights> European Commission </dc:rights>
</rdf:RDF>
```

Table of Contents

About this paper	6
Proposal for a profile that "uses"	8
Discussion	9
Related work	10
Bibliography	11

About this paper

This paper proposes for wider discussion and adoption some minimal guidelines for the construction of Application Profiles. An Application Profile declares the metadata model of an information service or application in a form directly usable in a Semantic Web environment – in other words, using a common grammar (RDF) and referencing only metadata terms that have formally been declared with namespace URIs.

After articulating three sets of assumptions, this paper proposes a simple model for profiles based on the verb "uses". This model was developed in a two-year project "SCHEMAS – A Forum for Metadata Schema Implementers", an Accompanying Measure of the EU's Information Society Technologies research programme. The full SCHEMAS model has been described in a technical paper [SCHEMAS, SCHEMAS-JODI]. This paper, in contrast, presents the core principles and assumptions underlying the SCHEMAS model as a basis for discussion in wider circles. The presentation of the "uses" model is followed by a discussion of design issues, potential uses of profiles, and related work.

ASSUMPTION 1

The metadata of the world is structured according to an uncontrolled diversity of data models, many of which are of a pragmatic, inconsistent nature. The large-scale integration of metadata, then, implies normalisation to a common grammar, even if the process of translation is imperfect .

This first assumption is in line with what one might call the "Semantic Web hypothesis". The W3C describes its vision of Semantic Web as "having data on the Web defined and linked in a way that it can be used for more effective discovery, automation, integration, and reuse across various applications" – both "by automated tools as well as by people". [SEMANTIC-WEB] Underlying the Semantic Web vision are the hypotheses that a shared grammar is needed to ensure that humans and software will interpret metadata consistently; that clusters of simple Subject-Predicate-Object statements in the style of Resource Description Framework (RDF) can describe most of the data processed by machines; and that more complex grammars would, at any rate, not interoperate in a massively diverse Web environment.

RDF statements use Uniform Resource Identifiers (URIs) to designate both the resources described – as well as the metadata terms used to describe them – with unambiguous identifiers that are unique points on a world-wide information space. These points, in turn, can be used as anchors for joining or merging multiple statements drawn or extracted from multiple sources. It is recognised that the process of normalising the diversity of metadata constructs of the world to a simple, uniform, almost pidgin-like statement grammar may involve a certain loss of specificity, and that exporting statements to unintended contexts may not always make sense, but these problems are accepted as an inevitable aspect of imperfect communication in an imperfect world. The assumption is that creating coherence – "making sense", of metadata on a grand scale must involve an imperfect process of translation, even simplification. Rather, the more modest goal is "partial understanding" – the lossy

and selective merging of data from underlying models that are semantically and structurally richer and more diverse.

To Tim Berners-Lee, the imperfect nature of this understanding is an inevitable limitation to the prospect of sharing data between software and resources that have been designed independently. Instead of expecting machines correctly to interpret the models people have designed, however, his vision of a Semantic Web involves asking people to make the extra effort to speak, as it were, machine-understandably. In this spirit, the model presented here provides a construct for implementers to make the extra effort of specifying a normalised projection of their metadata for merging with metadata from other sources on a large scale.

ASSUMPTION 2

The number of agencies declaring and defining “standard” metadata terms should be relatively small compared with the number of information providers “using” or “adapting” those standards for their metadata. In a Semantic Web environment, where metadata must be processable by machines, there is a need for good-practice schema constructs that make clear the distinction between “declaring” and “using”.

Until the nineteenth century, each library simply invented a local system for cataloguing the books under its roof, but as libraries began to pool access to their collections it became necessary to converge on regional and international standards. By analogy, early computerised databanks needed only to define their schemas internally until networks made it possible to merge access to multiple resources simultaneously. For many reasons, it is desirable now that information providers converge on metadata standards, and that these standards be manageable in number. Rather than reinventing common terms such as Title or Date, in other words, most providers should design their local schemas with appropriate terms from existing standards.

In an environment where metadata will increasingly be processed automatically, it is important that such standards be declared and published in a machine-understandable form, and promoting a data model for doing so has been a key focus of the W3C Semantic Web effort. Specific policies and guidelines for using that data model to declare schemas, however, fall to some extent outside the scope of W3C per se. This paper, for example, posits that the distinction between “declaring” metadata vocabularies as official or de-facto standards and “reusing” those vocabularies for particular applications is a key organising principle for a well-ordered metadata world. If RDF schemas along the lines of those published by the Dublin Core Metadata Initiative and by W3C itself suggest good-practice principles for “declaring” vocabularies, this paper offers a model for schemas that “reuse” such vocabularies.

ASSUMPTION 3

In order to meet the descriptive needs of particular applications, information providers often draw on multiple metadata standards, “mixing and matching” as needed and adapting or annotating standard

definitions with domain-specific guidelines and examples. In the interests of harmonisation, providers want to know how colleagues in related domains have done this. Such information should be harvestable directly from the information providers for delivery in term-level indexes of metadata semantics (“registries”).

In order to control the processing of metadata by software, many metadata applications define what one might call a “document validation schema”, such as an XML schema or Document Type Definition (DTD), to specify how the tag structure of specific metadata records is to be parsed and validated. Since XML itself places no particular restrictions on the nesting of elements, such tag structures can fulfill local, pragmatic needs yet be difficult to generalise. To cite one very common example, the attributes of an author – name, affiliation, email address, and fax number – are often nested within an author element. Such a structure can be processed using an XML schema to define the parent-child relation between, for example, Author and Fax. However, if the task is to merge metadata from a diversity of sources, it will be very difficult to relate their differently nested tag structures automatically, and heuristics for doing so in particular domains may not scale up to the Web as a whole.

A “semantic schema”, in contrast, is designed to show how metadata terms are defined and how those terms relate to terms defined in other such schemas. “Declarative” semantic schemas declare metadata terms and definitions for use in applications. A standard metadata vocabulary such as Dublin Core, for example, can be published as an RDF schema that assigns each element a namespace URI, specifies how that element is related to others, and defines encoding schemes for qualifying their potential values.

The Application Profile proposed in this paper is a particular type of semantic schema – one that limits itself to specifying how a particular metadata model “uses” terms defined in various declarative schemas. Upon such a declaration of usage can be hung various annotations to provide additional information about how those terms have been adapted or constrained for specialised purposes. By using RDF statements to associate annotations from a variety of sources with metadata terms defined uniquely in “declarative” schemas, Application Profiles can straightforwardly be integrated into RDF-based metadata registries – term-level indexes of declarative schemas and Application Profiles – for use in standardisation and harmonisation efforts. In a registry index, URIs serve as anchors for merging multiple references to specific metadata terms, allowing a range of useful queries for discovering which projects or services use which terms in which kinds of application contexts, together with which sorts of controlled vocabularies or value schemes.

Proposal for an Application Profile that "uses"

An Application Profile consists of a series of RDF statements centered around the verb "uses". The object of such a statement must be the namespace URI of a metadata term as defined in a declarative semantic schema (described above). The subject must be a URI representing the metadata that uses this metadata term. Loosely translated, then, an Application Profile makes statements such as “This metadata uses dc:title” or “This metadata uses foo:email”, where dc: and foo: resolve to the namespace URIs of terms defined in declarative semantic schemas.

With these simple “X uses Y” statements, in turn, can be associated any number of annotations with arbitrary types and amounts of further information, commentary, or technical documentation. For example, the statement might say, “This metadata uses foo:email to describe bar:authors” or “This metadata uses dc:title to describe bar:collections, defining it as ‘A name given to the collection’.” Interested readers will find some examples in a technical paper from the SCHEMAS Project. [SCHEMAS-JODI]

Discussion

- The verb “uses” is not currently part of any formal schema language. If it were now to be considered for wider use as a definitive feature of Application Profiles, it would be desirable to find an appropriate maintenance organisation and assign a stable namespace URI to this term.
- Declarations of terms "used" from namespaces can serve as hooks on which to hang any number of documentary annotations, such as locally relevant usage examples or restrictions on allowable metadata values. Such a profile can declare how "standard" metadata terms have been combined and optimised for a particular application, function, organisation, or user community.
- Registries should be able to harvest such profiles to construct an overview of metadata vocabularies and their use in particular domains. Helping implementers of information projects and services find out about metadata terms in use – both official definitions and local variations – will encourage harmonisation of metadata usage within particular fields and applications. By making metadata language "visible" to users, registries should facilitate the identification of empirical usage trends, feeding back into a more "bottom-up" process for making standards.
- Profiles must be declarable in a distributed manner. Managing a registry as a centralised database effort is neither scalable beyond a few dozen vocabularies nor sustainable as a project over the longer term. In general, profiles must be maintained by and directly harvested from the information providers themselves in a distributed, open Web environment.
- Profiles must be simple to understand and to create. Making information providers responsible for declaring their own profiles implies a pedagogical effort to clarify the purpose of a profile, along with easy-to-understand guidance materials and templates that automate the generation of XML encoding that can be harvested by registries.
- Profiles need not exhaustively describe underlying metadata models. In many cases, it may make more sense to provide excerpts. Not all parts of a local schema are necessarily important for interoperability, so not everything in a metadata model needs necessarily to be translated into an Application Profile.
- Metadata standards will evolve over time. By implication, the mappings contained in an application profile will also be subject to re-interpretation in order

to optimise, update, or correct the linkage of a local model to existing standards. For example, much of the information being made available on the Web is already described with metadata that was designed before the emergence of modern metadata standards and namespace URIs, or at any rate with local application needs, rather than interoperability on the Web, foremost in mind. The process of retrospectively mapping the terms of an existing local schema to terms in existing standards may therefore involve a process of interpretation. For example, a local XML tag called "TITLE" may be interpreted with hindsight as a Dublin Core title (<http://purl.org/dc/elements/1.1/title>).

- Potentially, profiles might serve as the basis for automated mapping or conversion between vocabularies. For example, a profile might serve as a template for exporting local metadata into a form that is easier to integrate with metadata from other sources.
- The metadata communities of the world have evolved a diversity of data models, to which they associate a diversity of metadata values, from simple strings to more structured entities. Further research will be needed into the nature and consequences of error and loss when merging values associated with such disparate models within a normalised Semantic Web framework.

Related work

Numerous standardisation communities have developed some notion of “profile”:

- The Z39.50 community uses "profiles" for constraining potential options and parameter values, where left open by standards specifications, to those required by a particular application (such as GILS or WAIS), function (such as simple author-title-subject searching), or user group (such as chemists or musicians). According to the "Framework and Taxonomy of International Standardized Profiles" (ISO TR 10000), a profile specifies how standards, particularly protocols, can be used in combination for meeting such requirements. [Z3950]
- The Dublin Core Metadata Initiative has a loosely defined notion of "application profile" for describing how a core element set is extended for simple description in specific domains, such as Education and Government Information. [DCMI]
- In IEEE standardisation committees for learning technology, a "standards profile" is "a technique of referencing (in contrast to defining) technical specifications... [permitting] the creation of a bundle of standards, each one tailored, extended, or constrained to meet the needs of the committee developing a standards profile. The point of using standards profiles is to reuse existing standards wording without having to recreate the words". [IEEE1484]
- To users of the Digital Object Identifier, a DOI Application Profile is "the functional specification of an application (or set of applications) of the DOI System to a class of intellectual property entities that share a common set of attributes" for the purpose of enabling particular applications, from simple resource discovery to complex rights management. [DOI-AP]

- The Federal Geographic Data Committee distinguishes between its the Content Standard for Digital Geospatial Metadata, and a profile based on that standard, which "describes the application of the Standard to a specific user community". A profile "always contains the Standard, plus modifications to the optionality or repeatability of non-mandatory elements in the Standard" and "may also contain extended elements"; it may be formalised through the FGDC process or used informally by a user community. [FGDC]
- ISO/DIS 19115, another standard for geographic datasets, likewise provides for the development of "community profiles" within user communities, nations, or organisations. [ISOTC211]

In addition, Jane Hunter has reported that "TV-Anytime, MPEG-21, and the Open Archives Initiative are demanding application profiles which combine elements from a number of different existing standardised metadata schemas whilst maintaining interoperability and satisfying their own specific requirements through refinements, extensions and additions." [HUNTER]

The problem of mapping diverse conceptual structures is not unique to the metadata community -- the same problem is being addressed in other contexts for mapping databases, thesauri, and ontologies. Ontologies, for example, may be expressed in different representational languages on the basis of ill-defined or inconsistent models which may be pragmatically or locally useful, but difficult to relate automatically.

From their various perspectives, these diverse communities seem to be reaching the same conclusion: that the mapping process can be automated only in part, and that manual intervention by experts is usually needed to complete (or correct) the job. This implies the more general conclusion that the large-scale merging of metadata cannot reliably be left entirely to algorithms and heuristics, but would benefit from mapping constructs – such as Application Profiles of the type discussed here – and from good-practice guidelines for encoding and making those profiles available for use by registries.

Bibliography

- [DCMI] <http://dublincore.org>
[DCMI-NAMESPACE] <http://dublincore.org/documents/2001/09/17/dcmi-namespace/>
[DCMI-REGISTRY] <http://www.dublincore.org/groups/registry>
[DOI-AP] <http://www.doi.org/doi-ap.html>
[DESIRE] <http://desire.ukoln.ac.uk/registry/>
[EOR] <http://eor.dublincore.org>
[FGDC] <http://www.fgdc.gov/metadata/csdgm/profile.html>
[HEERY] <http://www.ariadne.ac.uk/issue25/app-profiles/intro.html>
[HUNTER] <http://www10.org/cdrom/papers/572/index.html>
[IEEE1484] <http://edutool.com/pmp/>
[ISOTC211] <http://www.statkart.no/isotc211/scope.htm>
[SCHEMAS] <http://www.schemas-forum.org/>
[SCHEMAS-FRAMEWORK] <http://www.schemas-forum.org/stds-framework/>
[SCHEMAS-JODI] <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/>

[SCHEMAS-WATCH] <http://www.schemas-forum.org/metadata-watch/>
[SCHEMAS-WORKSHOPS] <http://www.schemas-forum.org/workshops/>
[RDFCore] <http://www.w3.org/2001/sw/RDFCore/>
[RDF-SCHEMA] <http://www.w3.org/TR/rdf-schema/>
[SEMANTIC-WEB] <http://www.w3.org/2001/sw/Activity>
[XML-SCHEMA] <http://www.w3.org/TR/xmlschema-0/>
[Z3950] <http://leweb.loc.gov/z3950/agency/profiles/about.html>