

Metadata Application Profiles

Rachel Heery, UKOLN, University of Bath

www.ukoln.ac.uk

Robina Clayphan, British Library

www.bl.uk

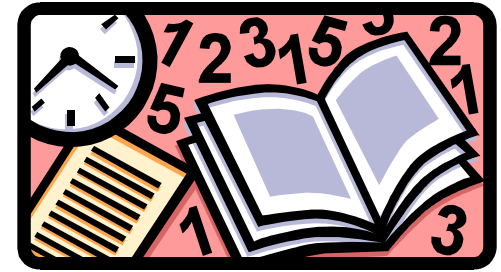
DC-2005: International Conference on Dublin Core and Metadata Applications, University Carlos III of Madrid.

Tutorial 5: Application Profiles (part I)

15 September 2005



Tutorial will answer.....



What is an application profile?

Why do we need application profiles?

Why should I declare an application profiles?

How can I build an application profile?

What is best practice for documenting an application profile?

And what is happening in the 'real world'?

Case study 1 : British Library Application Profile

Can I find out more? (yes, there is a reading list!)

DC Application Profile definition

- Specifies which metadata terms an organization, information provider, or user community uses in its metadata
- Identifies the terms used to describe a resource
- Optionally provides additional information about term usage e.g. how encoding schemes constrain values



see DCAP Guidelines, CEN Workshop Agreement 14855:2003

More definitions...

- **Term:** The generic name for a *property* (i.e. *element* or *element refinement*), *vocabulary encoding scheme*, *syntax encoding scheme* or concept taken from a controlled vocabulary (concept space).
- A **property** is a specific aspect, characteristic, attribute, or relation used to describe *resources*.
- Within DCMI, **element** is typically used as a synonym for *property*.
- A **vocabulary encoding scheme** is a *class* that indicates that the *value* of a *property* is taken from a controlled vocabulary (or concept-space), such as the Library of Congress Subject Headings.



Context

- Evolution of DCMI metadata vocabulary
 - 10 years old!
- Experience after 10 years of implementation
 - Informing practice for extensions and customisation
- Connected work on schemas and registries
 - DESIRE, SCHEMAS, CORES, MEG Registry
- MMI-DC workshop within the European Committee for Standardisation (CEN)
 - CEN Working Agreements
 - CWA 14855:2003, CWA 15248:2005



Why do we need application profiles?

Why should I declare an application profile?



Proliferation of metadata

Metadata is required for:

- Resource discovery
 - Enterprise portals
 - Subject gateways
 - Metasearch
- Managing high throughput eScience
- Expressing Digital Rights
- Ensuring preservation

Appropriate terms must be identified wherever metadata is needed



Proliferation of metadata

Increase in metadata vocabularies

- DCMI
- IEEE LOM
- MODS
- MARC 21
- UNIMARC
- MPEG-7

- FOAF
- RSS



Implementor perspective

Implementors are seeking a metadata vocabulary for their particular service or system

- Implementors approve of re-use
- Implementors acknowledge importance of interoperability
- but there is pressure to satisfy local requirements and to be innovative

Tension between using standard terms in a vocabulary and localisation



Proliferation of localised extensions

- Metadata standards vocabularies are published
but
- Implementor adaptations and extensions are not made widely available

Sharing semantics will reduce duplication and repetition

We need to exchange information about metadata terms we use

What terms do your metadata descriptions use?

A DCAP expresses in a structured way

- Which standard terms are used in an application
- Source of terms
- Usage constraints



Benefits of documenting terms we use

- To provide authoritative specification of term usage
- To facilitate interoperability by informing potential users
- To support evolution of vocabulary
- To encourage alignment
- To enable interpretation of legacy metadata



Profiling 'standards' is not new...

- Z39.50 application profiles
 - sub-sets of standard appropriate for application area
- IEEE LOM
 - UK Common Metadata Format
- METS profiles
- and so on...



Extensibility of metadata vocabularies is not new...

- Warwick Metadata Workshop, 1996
- MARC
 - 9XX local tags
- PRISM (Publishing Requirements for Industry Standard Metadata)
- and so on...



How do I build an application profile?

...need to meet service and systems requirements for metadata...



Reviewing requirements for metadata

- Analyse functionality required (...by means of use cases)
- Will an existing standard vocabulary meet the requirements?
- If not what extensions are needed?

For comparing metadata and functionality see TEL matrix

TEL metadata functionality matrix

Score whether:

- element contributes to a function
- element is required or contributes to a large extent
- element's presence may trigger the portal to offer a function

Functionality criteria:

- Resource discovery
- Identification
- Multilinguality
- Authority service
- Administration
- Authorisation
- Copy cataloguing
- And so on

Britta Woldering 'The European Library: Integrated access to the national libraries of Europe' January 2004 , Ariadne Issue 38.
<http://www.ariadne.ac.uk/issue38/woldering/intro.html>



Making choices between metadata vocabularies

- Look at what others in the domain are using
- Consider:
 - stability/volatility of the standard (and whether it really IS a standard)
 - how the community for the standard integrates new needs and ideas
 - startup and maintenance costs for use in an individual project (higher for more complex formats and implementations)
- Document choices and reasoning for your successors (they will thank you)
(acknowledgements to Diane Hillmann!)



What terms to use to meet your requirements?

Decision tree:

- Use existing DCMI term whenever possible
- Can requirement be met with a new encoding scheme value, or a new encoding scheme?
- Are there suitable terms already used and declared in other DC application profiles?
- If not declare a new local property....

Examples of DC Application Profiles

- JISC IE Services Registry AP

<http://iesr.ac.uk/profile/>

- TEL AP (The European Library)

http://www.europeanlibrary.org/tel_application_profile_v.htm

- DCMi APs

- Library AP

<http://dublincore.org/documents/2004/09/10/library-application-profile/>

- Collection Description AP

<http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>

- Education AP

http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile



The European Library (TEL) Application Profile

- Starting point was DC-Lib AP
- TEL-specific additions to support desired functionality e.g.
 - OpenURL (get local services for this record)
 - RecordId (get original record)
 - Thumbnail (thumbnail image)
- Why? acknowledged need for controlled evolution of metadata terms
 - the ability to add future functionality may depend on additional terms
 - new sectors/collections may require specific terms



http://www.europeanlibrary.org/tel_application_profile_v.htm

Declaring an application profile....



What does an application profile express?

Implementors need to declare various characteristics of their schema

- Terms in use
- Whether a term is mandatory
- Any particular usages
- Permitted encoding schemes for values
- Other rules for content

DCAPs: human readable

First step is to address need for human readable DCAP

Drawing on:

- CEN guidelines
- DC WG practice
- DC Usage Board advice
- DC Abstract Model



CEN Workshop Agreement 14855

Dublin Core Application Profile Guidelines.

Thomas Baker, Makx Dekkers, Thomas Fischer, Rachel
Heery

26 September 2003

<http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/cwa/cwa14855.asp>



DCAPs : machine readable

Ongoing work...

- CWA 15248 suggests a machine-processable representation using conventions of RDF
- Would support automated services such as data exchange, query
- Subject of ongoing experimentation and research
- Need to reach consensus on a DCAP data model



Caveat...DCAPs are in transition

- Differences in terminology and attributes from CWA 14855 to CWA 15248
- As DC Abstract Model becomes 'embedded' further adjustments will occur
- Things change



What is good practice for
documenting application profiles?

..based on CWA 14855...



Format of (human readable) DCAPs

- Normalized and readable view of Dublin Core based schemas for use by humans
 - No particular format mandated: plain text, Web pages, Powerpoint...
- Enough structure for future conversion into machine-processable expressions (eg, RDF)
 - Future conversion not assumed to be automatic
 - Caveat: normalized documentation does not in itself address deeper problems of interoperability between metadata models.



DCAPs for a class...

- Declares the property usages in a 'class' of metadata descriptions
 - e.g. DC Collection Description application profile describes the properties and encoding schemes (terms) used in a description of a collection
 - TEL application profile describes properties and encoding schemes used in a description of a European national libraries resource
 - DC education application profile describes properties and encoding schemes used in a description of an educational resource.

DCAPs: Good practice

- Follow DC Abstract Model (terms should adhere to definitions of properties, sub-properties, encoding schemes etc)
- Should consist of Descriptive Header and Term Usages
 - Descriptive Header
 - DC-based description
 - Optional Preamble
 - Term Usage
 - Terms used identified with "appropriate precision"
 - May be annotated with additional attributes and constraints



Descriptive header

Should include:

- Description of the AP using Dublin Core
- Title
- Contributor
- Date
- Identifier
- Description (explaining context in which DCAP is intended to be used)

Descriptive headers

Mandatory Header

- Description of the AP using Dublin Core
- Title
- Contributor
- Date
- Identifier
- Description
 - explaining context in which DCAP is intended to be used

Optional preamble

- Technical or formatting conventions
 - Documenting namespace prefixes
- Citing web pages where terms used are defined
 - e.g. <http://www.dublincore.org/documents/dcmi-terms>



Attributes of Term Usages

- *Identifying attributes*
 - **Term URI, Name, Label, Defined By**
- *Definitional attributes*
 - **Definition, Comments, Type of Term**
- *Relational attributes*
 - **Refines, Refined By, Encoding Scheme For, Uses Encoding Scheme, Similat To**
- *Constraints*
 - **Obligation, Condition, Datatype, Occurrence**

Principle of Appropriate Identification

Terms should be identified as precisely as possible ("appropriate precision")

- URIs should be used when available (see CORES Resolution)
- Terms to which URIs have not (or not yet) been assigned should be identified using other attributes as appropriate

Declaring terms

- Preferred: cite term's URI if available
 - Term URI <http://purl.org/dc/terms/audience>
- Or if a term has been declared somewhere, cite the defining document and its name
 - Name **attendancePattern**
 - Label **Attendance Pattern**
 - Defined By <http://someones-project.org/schema.html>
- If term has not been declared elsewhere, *Defined By* should cite the DCAP itself
 - Name **starRatings**
 - Label **Star Ratings**
 - Defined By <http://myproject.org/profile.html>



Principle of Readability

“include enough information in Term Usages to be of optimal usefulness for the intended audience”

- Even if this includes redundant information which, in a machine-processable schema, might be fetched dynamically from another source
- Order of attributes may be changed for readability (though it may make visual comparison harder)
- Unused attributes can simply be omitted from display



Readability of Term Usages

Principle of Readability allows flexibility in presentational style

- Redundant attributes do not need to be displayed (as blank)
- Order of attributes may be altered for visual effect (not significant for future machine-processable representations)
- DCAP may want to group terms by Type of Term
- Attributes should be repeated as necessary



Declaring controlled vocabulary terms

- Generally not the role of DCAPs to declare controlled vocabularies of values
- Ideally, should be declared in separately citable documents external to a DCAP
- However, short lists of possible values may be documented in a Comment field

Declaring Encoding Schemes

Options

- Can be declared one-by-one in the Term Usage of an Element in the field "Has Encoding Scheme"
- Field "Has Encoding Scheme" can point to a list of encoding schemes somewhere (e.g. "use RDN Subject Encoding Schemes...")
- If Encoding Schemes need to be annotated, a separate Term Usage may be created for each



Cite URIs of terms

- URIs are (ideally) unique and unambiguous:
 - Example: <http://purl.org/dc/elements/1.1/title>
- For readability Qualified Names can be cited in Name field
 - Example: [dc:title](#)
 - Explain in the Preamble that this is the case
 - Cite the URIs



Example: JISC IE Service Registry Application Profile - Namespaces

- **Term URI**
 - <http://purl.org/dc/elements/1.1/>
 - **Name**
 - dc:
 - **Label**
 - [Dublin Core](#)

 - **Term URI**
 - <http://purl.org/dc/terms/>
 - **Name**
 - dcterms:
 - **Label**
 - Dublin Core terms

 - **Term URI**
 - <http://purl.org/rslp/terms#>
 - **Name**
 - rslpcd:
 - **Label**
 - [RSLP Collection Description](#)
- **Term URI**
 - <http://purl.org/cld/colldesc/type/>
 - **Name**
 - colldesctype:
 - **Label**
 - [NISO Metasearch Initiative Collection Description Type Vocabulary](#)

 - **Term URI**
 - <http://iesr.ac.uk/terms/#>
 - **Name**
 - iesr:
 - **Label**
 - [JISC Information Environment Service Registry](#)



Examples of well structured DCAPs

- RDN OAI application profile
http://www.rdn.ac.uk/oai/rdn_dc/
- Renardus Application Profile
<http://renardus.sub.uni-goettingen.de/renap/renap.html>
- JISC IE Services Registry Application Profile
<http://iesr.ac.uk/profile/>
- DC Collection Description Application profile
<http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/>



RDN OAI Application Profile - header

Title	RDN OAI Application Profile
Contributor	Andy Powell
Date	2003-03-23
Identifier	
Description	<p>This document expresses the application profile established by the Resource Discovery Network (RDN) to be used by RDN partners for harvesting of records using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). The Application Profile is expressed according to guidelines published by the CEN/ISSS [Reference]. Full user documentation for the Application Profile, together with associated XML schemas, is available at http://www.rdn.ac.uk/oai/rdn_dc/</p>



RDN OAI Application Profile – term usage

Name	Subject
Term URI	http://purl.org/dc/elements/1.1/subject
Has Encoding Scheme	DC Subject Encoding Schemes
Has Encoding Scheme	RDN Subject Encoding Schemes
Comment	RDN Subject Encoding Schemes are available from http://www.rdn.ac.uk/publications/cat-guide/subject-schemes/
Obligation	Recommended

IE Service Registry application profile - header

dc:title	JISC Information Environment Service Registry (IESR) Application Profile
dc:creator	Ann Apps
dc: date	2005-09-06
dc:identifiwr	http://iesr.ac.uk/profile/
dc:description	An application profile for the JISC Information Environment Service Registry (IESR), described according to the Dublin Core Application Profile Guidelines (CEN/ISSS CWA14855).
dc:rights	This work is licensed under a Creative Commons Licence : <i>Attribution Required; Non-Commercial; Share-Alike</i>
version	http://iesr.ac.uk/profile/2005/09/06/
status	Implemented



IE Service Registry application profile - term usage

Term URI	http://purl.org/rslp/terms#locator
Name	Locator
Definition	The URI of the access point for the service
Comments	For Z39.50 services this will be a URI beginning z3950s
Has encoding scheme	http://purl.org/dc/terms/URI
DataType	<URI>
Occurrence	Min 1;Max 1
Searchable	1209, 1016, 1017 etc

Further reading

CEN Workshop Agreement 14855: Dublin Core Application Profile Guidelines. Available at:

<http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/s/cwa/cwa14855.asp>

CEN Workshop Agreement 15248: Guidelines for machine-processable representation of Dublin Core Application Profiles. Available at:

<http://www.cenorm.be/cenorm/businessdomains/businessdomains/iss/s/cwa/cwa+15248.asp>

Thomas Baker, Makx Dekkers, Rachel Heery, Manjula Patel, Gauri Salokhe, *What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives.* Journal of Digital Information, Vol.2, no. 2, November 2001. <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/>

Thomas Baker and Makx Dekkers, *Identifying metadata elements with URIs: the CORES Resolution.* D-Lib magazine, July/August 2003.

<http://www.dlib.org/dlib/july03/baker/07baker.html>



Acknowledgements to colleagues, in particular Tom Baker and Diane Hillmann.

Many thanks to Pete Johnston for his continuing efforts to formalise application profiles.



Questions?



Case study....



DCMI: Properties and Values

Resource URI	Statement	
ex:book1	Property URI	Value string (or Value URI)
	dc:subject	Semantic Web

see Pete Johnston 'Element Refinement in Dublin Core Metadata' June 2005.

<http://dublincore.org/documents/2005/06/13/dc-elem-refine/>

