

## About this draft review

---

This document is a 'trial run' for a full review of the CD Application Profile by the Usage Board. The draft is work-in-progress; it is intended for internal use of the Usage Board and should not be cited or quoted outside that context. The purpose of this document is to help the Usage Board visualize the end result of the review process -- its outline and (roughly) its length and granularity of detail.

## Usage Board Review of the Collection Description Application Profile -- Draft

---

The DCMI Usage Board undertook an assessment of the Collection Description Application Profile (CDAP) at a meeting in Manzanillo Mexico [1] on Sunday, 1 October 2006. Members present were Tom Baker (chair), Diane Hillmann, Andy Powell, Akira Miyazawa, Stuart Sutton, Joe Tennis, Andrew Wilson (designated shepherd of the Collection Description proposal), and Makx Dekkers (ex officio). This note holds an interim response to the DCMI Collection Description Working Group, short of being a formal review of a submitted application profile. This note includes comments on the Collection Description application profile as a whole. See [1] for links to the actual documents reviewed.

== About the Collection Description Application Profile ==

*The following sections in italics are pasted from [2] and should be edited down to summaries.*

The purpose of the CD Application Profile is to "provide a means of creating simple descriptions of collections and collection-descriptions suitable for a broad range of collections. It is designed primarily to support the discovery and selection of collections, though it may be used to support other functions such as collection management too. It is not intended to describe every possible characteristic of every type of collection".

### Collections and collection-level description

---

The term "collection" can be applied to any aggregation of physical or digital items. Those items may be of any type, so examples might include aggregations of natural objects, created objects, "born-digital" items, digital surrogates of physical items, and the catalogues of such collections (as aggregations of metadata records). The criteria for aggregation may vary: e.g. by location, by type or form of the items, by provenance of the items, by source or ownership, and so on. Collections may contain any number of items and may have varying levels of permanence.

A "collection-level description" provides a description of the collection as a unit: the resource described by a collection-level description is the collection, rather than the individual items within that collection. Collection-level descriptions are referred to in Michael Heaney's *An Analytical Model of Collections and their Catalogues* as "unitary finding-aids" [AMCC].

Collection-level description enables a collection provider to

- disclose information about the content and availability of collections to users where item-level metadata does not exist or is not available, or where the provision of item-level detail is not required or appropriate

It enables a user to

- discover and locate collections of interest
- select collections to explore on the basis of a summary description
- compare collections as broadly similar objects, even where items (and/or item-level metadata) are heterogeneous
- understand conditions of access and use
- interpret collections (and items within collections)

Increasingly, many of these functions - notably the discovery, location, selection and comparison of collections - are being carried out by software acting on behalf of a human user, perhaps in accordance with user preferences or with parameters describing the scope of a particular service.

## **Functional Requirements of the DC CD AP**

---

The DC CD AP is intended to provide a means of creating simple descriptions of collections and collection-descriptions suitable for a broad range of collections. It is designed primarily to support the discovery and selection of collections, though it may be used to support other functions such as collection management too. It is not intended to describe every possible characteristic of every type of collection.

The aim is that the DC CD AP should support:

- the discovery of collections and collection-descriptions of potential interest, by enabling searching on various attributes of the collection including:
  - the name or title of the collection or collection-description
  - the subject and coverage of the collection
  - the nature or genre of the items within the collection or collection-description
  - the media type or format of the items within the collection or collection-description
  - the entity that created the collection or collection-description
  - the entity that owns the collection or collection-description
  - relationships between collections, between collection-descriptions and between collections and collection-descriptions
  - the identification of a known collection or collection-description, by enabling the capture and disclosure of identifying attributes such as
    - the formal identifier(s) the collection or collection-description
    - the name or title of the collection or collection-description
    - a textual description of the collection or collection-description
  - the selection of one or more collections or collection-descriptions from amongst a number of discovered collections or collection-descriptions, by enabling the capture and disclosure of attributes such as
    - a textual description of the collection or collection-description
    - a description of rights held in/over the collection or collection-description and conditions of access and use
    - a description of the custodial history of the collection or collection-description
    - a description of the way in which items are added to the collection or collection-description
  - the identification of the location of the collection or collection-description

- the identification of the services that provide access to the collection or collection-description

## **CDAP Application Model**

---

The DC CD AP is based on a data model which is derived from that described in Michael Heaney's An Analytical Model of Collections and their Catalogues [AMCC] and Users and Information Resources: An Extension of the Analytical Model of Collections and their Catalogues into Usage and Transactions [EAMCC]. The model used here is both a subset and a simplification of that model. The entity type which in that model is referred to as "Resource Mediator" is referred to here as "Service".

## **Review of CDAP by the Usage Board**

---

The Board is of the view that this application profile is almost ready to be submitted as a complete proposal to the Usage Board for formal review and compliments the Working Group and the chair Pete Johnston, on the dedication and effort that has been put into the work. The Board asks the working group to resubmit its proposal for re-consideration and recommends the following changes and additions:

1. The data model used for the CDAP needs a better explanation. This explanation should include a statement about how the model in the Application Profile diverges from or otherwise amends the AMCC model cited as the basis for the profile. The Board feels that more explanatory material is needed to assist in relating the data model to the AP where dependencies exist.
2. The document should include a summary statement of the purpose and scope of the CDAP.
3. The Board feels that the distinction between "use of CDAP for collection descriptions" and "use of CDAP for collections of collection descriptions" is a bit confusing, and the CDAP should more clearly separate the two. This might be achieved by splitting the current CDAP into two separate APs -- one for collection descriptions and one for collections of collection descriptions. Editorial decisions of this nature are at the discretion of the working group.
4. The Board suggests the Working Group reconsider the change of label for the term 'Collection-Description' [2]. The Board feels that the old label, 'Catalogue or collection description' [3] is clearer and easier to understand because the label itself is an example of a typical use for the term.
5. The term 'Content' is not defined in the AP, but is used in a definition [4]. The Board asks the Working Group to include a definition of `content` in the Glossary. The same applies to some terms used in the CDAP that are defined in the DCMI Abstract Model, such as 'Rich Description'.

In general, the Board prefers that Application Profiles be able to 'stand alone' as documents; ideally, readers of the CDAP should not need to refer to external documents in order to understand its logic and suggests that the working group consider adding some additional contextual information from external documents as an aid to the reader.

## **Documentation Submitted**

---

*This list to be turned into a more narrative characterization of documentation submitted.*

- Full CDAP profile <http://dublincore.org/groups/collections/collection-application-profile/2006-08-24/>
  - The main change in this version is to separate out a list of properties used to describe a

Collection that is also a Collection-Description, as discussed recently.

- 1. Introduction. The Introduction describes the purpose of describing collection level resources, sets out the functional requirements for the application profile, and describes the data model used as the basis for the profile -- the Analytical Model of Collections and their Catalogues developed by Michael Heaney.
- 2. Vocabularies/Namespaces Used in this DCAP. Section 2 lists namespaces used in CDAP: Dublin Core Metadata Element Set, v1.1, Dublin Core Metadata Terms, Dublin Core Type Vocabulary, MARC Relator Code Properties, Collection Description Terms, Collection Description Type Vocabulary Terms.
- 3. Describing a Collection. Starts out defining the class `Collection`. This is followed by documentation on of thirty properties which can be used to describe a Collection.
- 4. Describing a Collection-Description. Defines the Collection Description class and then provides definitions for the twenty-six properties which can be used to describe Collection Description. The 26 terms include 4 terms not appearing in the list above and exclude 8 terms from the above list.
- 5. Vocabulary Encoding Schemes Used
- 6. Syntax Encoding Schemes Used
- 7. Administrative Metadata
- Summary CDAP profile <http://dublincore.org/groups/collections/collection-ap-summary/2006-08-24/>
  - To minimise redundancy, I moved most of the introductory material out of the summary document and expanded the introduction in the main DCAP document. That really means that the summary document is pretty much a "ready reference" tool only, and readers coming to the DC CD AP really need to look at the full document to understand it.
  - Intended to be a very condensed view of the former -- to enable people to see "at a glance" what terms are used in the DC CD AP (which I think is quite difficult to get from the full doc in its current format) - and/or to act as a sort of "rich table of contents" for the former, but it is not intended to be stand-alone. I deliberately stripped out some of the information that was previously in the summary to avoid duplication/redundancy.
- Documentation submitted by Collection Description WG:
  - I've also created a separate document describing the "Collection Description Terms", i.e. the new properties, vocabulary encoding schemes and syntax encoding schemes coined for use in the DC CD AP:
  - Dublin Core Collection Description Terms  
<http://dublincore.org/groups/collections/collection-terms/2006-08-24/>
  - Dublin Core Collection Description Type (CDType) Vocabulary  
<http://dublincore.org/groups/collections/colldesc-type/2006-08-24/>
    - The Collection Description Type Vocabulary has been updated, to correct errors and to separate the descriptive text out into a one-line "Definition" and a more discursive "Comment".
- There are full descriptions of the sets of terms in three vocabulary encoding schemes:
  - Collection Description Frequency [a vocabulary encoding scheme]  
<http://dublincore.org/groups/collections/frequency/2006-08-24/>
  - Collection Description Accrual Method [a vocabulary encoding scheme]  
<http://dublincore.org/groups/collections/accrual-method/2006-08-24/>
  - Collection Description Accrual Policy [a vocabulary encoding scheme]

<http://dublincore.org/groups/collections/accrual-policy/2006-08-24/>

[1] <http://dublincore.org/usage/meetings/2006/09/manzanillo/profile-cdap/html/>

[2] <http://dublincore.org/groups/collections/collection-terms/2006-08-24/#cldcollectionDescription>

[3] <http://www.ukoln.ac.uk/metadata/dcmi/collection-application-profile/2005-08-25/#dcdescription>

[4] <http://dublincore.org/groups/collections/collection-terms/2006-08-24/#clditemType>

Note - links to current documents about Application Profiles at:

<http://dublincore.org/usage/meetings/2006/04/profile-review/>

Pete on the RDF representations - 2006-08-29

-- <http://dublincore.org/groups/collections/collection-terms/2006-08-24/cldterms.rdf> --

<http://dublincore.org/groups/collections/colldesc-type/2006-08-24/cdtype.rdf> --

<http://dublincore.org/groups/collections/frequency/2006-08-24/freq.rdf> --

<http://dublincore.org/groups/collections/accrual-method/2006-08-24/accmeth.rdf> --

<http://dublincore.org/groups/collections/accrual-policy/2006-08-24/accpol.rdf>

The RDF data should essentially be an alternative representation of what is in the HTML docs i.e. the only question marks are to do with that representation (e.g. the relationship between a value in a VES and the VES, if we change the DCAM so that it is not is-instance-of (rdf:type))

I put up RDF/XML representations of the "collection description terms", the type vocabulary, and these three vocabularies, and the PURLs for the terms should de-reference to those RDF/XML docs (i.e. in the same way DCMI serves one doc per "namespace"). But some of that data is incomplete/tentative, pending some decisions about the DCAM and/or property ranges/domains, so at the moment it's probably best to treat the RDF/XML stuff more or less as a "placeholder".

All the term URIs (I hope) de-reference to something useful -- at the moment that's an RDF/XML document, but in the future we might set things up so that agents can get alternative representations (e.g. HTML for a browser displaying stuff to a human reader, RDF/XML for an app that wants to get the data about the relationships between terms in a form it can act on).

Note that the current RDF representations listed above are tentative/incomplete, pending discussions in the DC Architecture WG about changes to the DCMI Abstract Model, which would have an impact on how e.g. we describe the relationship between a vocabulary encoding scheme and a member term/value within that vocabulary encoding scheme.