



Dublin Core metadata schema: ADS interpretation and mapping for the ARENA project partners

Tony Austin [final] 13 Mar 2003

There are problems associated with Dublin Core (DC) in that different communities and even individuals within communities may interpret the semantics of specific elements differently; however, as long as the partners within ARENA use the same interpretation interoperability should be sustainable. Other problems concern granularity in that information may be held at different levels, for example, record and collection level. For the ADS it has become apparent that some DC elements are more easily dealt with at the record level and others at collection level. This may not be a problem for the ARENA partners. The ADS interpretation and mapping of its data to DC has been driven by a number of influences; notably *Discovering Online Resources Across the Humanities: A Practical Implementation of the Dublin Core* (Miller and Greenstein 1997), the Dublin Core Metadata Initiative (DCMI) definitions of elements (see DCMES, 2003) and more recently the UK *e-Government Metadata Standard* (2003).

The core DC schema contains 15 elements all of which are repeatable (although the ADS restrict DC_Identifier to a single unique identifier). DC elements can be qualified to form sub elements although the DCMI has yet to formally ratify any such extended schema; however, agreement within communities and groups can facilitate interoperability and the meaning of data items (e.g. a date such as 2003 is meaningless without a 'date_type' qualifier – thus 'record created' + '2003'). ADS mapping and examples are detailed below. It is not inclusive in that not all elements and sub-elements have to be mapped. Partners who are going to use OAI will need to map to qualified DC to facilitate import into ADS systems. Z targets can provide search results as unqualified DC but as noted above and as will become even more apparent below this is often provides meaningless information; however, some degree of qualification can be built into the database queries that generate a Z39.50 results set.

DC_Coverage is concerned with the spatial and temporal characteristics of the intellectual content of a resource including spatial coordinates and thus a really important element for people concerned with studying the Historic Environment.

examples

unqualified

Coverage

St Leonard's Churchyard; Eynsham Abbey

Eynsham

West Oxfordshire

Oxfordshire

England

1d 22' 20" W 51d 46' 42" N (stored in numeric form also)

Bronze Age

Early Medieval

Medieval

qualified

Hierarchical locative information is usually further qualified by a 'location type'; in England (they vary through UK) terms such as place or site name, parish, unitary authority, etc). Thus we get

Coverage.Location.Type	Coverage.Location.Description
PLACE	St Leonard's Churchyard; Eynsham Abbey
PARISH	Eynsham
UNITARY	West Oxfordshire
COUNTY	Oxfordshire
COUNTRY	England

Temporal attributes

Coverage.Temporal

Bronze Age

Early Medieval

Medieval

Coordinate information will probably be stored in a number of forms. For instance ADS store as OSGB grid references, numeric equivalent of these, longitude and latitude as degrees, minutes and seconds and longitude and latitude decimal equivalents. Arena partners will only be interested in the latter as a common shared coordinate system.

Coverage.Coordinate

1d 22' 20" W 51d 46' 42" N

(ADS use long lat decimal for spatial searching but display in results sets as degrees, minutes, seconds)

DC_Contributor

An entity responsible for making contributions to the content of the resource. This element is not widely used with a tendency to group this 'who' type information under DC element Creator qualified by Role. The ADS take this position.

DC_Creator

People or organisations responsible for creating intellectual content of a resource. In combining with DC_Contributor it covers both major and minor input. Effectively becomes 'who' is associated with a resource.

unqualified

Creator
Oxford Archaeological Unit
Oxfordshire Museums/NMRE
Oxfordshire Museums

qualified

Can be qualified by a 'role' sub-element which make clear what input is/was. As such can hold information about archive locations and curation therein.

Creator	Creator.Role
Oxford Archaeological Unit	Responsible for work
Oxfordshire Museums/NMRE	Paper/microfilm archive: location
Oxfordshire Museums	Artefact archive: location

Other roles might be 'responsible for dating' (e.g. a C14 lab) and so forth.

DC_Date

Dates associated with the creation and availability of a resource (i.e. its lifecycle) and thus can be project dates for an excavation or record creation and modification dates as examples. ADS conforms to ISO 8601 format of YYYY or YYYY-MM-DD for storage but present in more human readable forms

unqualified

Date
10 Sep 1999
1989
1992

qualified

Clearly date is not very useful without qualification of a 'date type'

Date.Type	Date
Project Start	1989
Project End	1992
Record Created	10 Sep 1992
Record Last Modified	

DC_Description

A free text description of the content of a resource

unqualified

Full excavation of part of the precinct of the medieval abbey, revealing part of the cloister and south ranges, kitchens and domestic buildings. Evidence of earlier activity included the 11th century Anglo-Saxon abbey and earlier minster, and a Bronze Age enclosure.

qualified

Would need qualification if multiple descriptions were to be used as, for example, for different audiences. ADS is not currently set up to deal with such a scenario

DC_Format

The physical or digital manifestation of the resource. Can be used for physical dimensions of an object or media types associated with a resource. ADS do not currently use but project partners may want to disseminate this sort of information. Various schema are recommended for controlling the use of format including

Internet Media Types (IMT) <http://www.isi.edu/in-notes/iana/assignments/media-types/media-types>

Système International d'Unités an internationally agreed standard concerning units of measurement at http://www.bipm.fr/enus/3_SI/si.html

DC_Identifier

String or number used to uniquely identify a resource.

unqualified

identifier

OAU_EEA_89-92-1

No need for qualification

In holding multiple collections ADS ensure uniqueness by adding a prefix to a depositor's internal identifier. For example a record within the Greater London SMR with an internal identifier of 123456 is assigned an auto-generated identifier of something like GLSMR-123456. This process will be applied to OAI harvested data but not Z39.50 targets who will retain Ids as they are.

DC_Language

Language of the intellectual content of a resource. Currently conformance to the ISO 639 two letter language codes is recommended by the DCMI

Unqualified

Language

en
ro
pl

No need for qualification

This can generally be auto-generated if languages are consistent (as for SMR type data)

DC_Publisher

Entity responsible for making a resource available in its present form. Thus names of partner organisations. ADS hold detailed publisher information for ArchSearch but currently only delivers publisher name through HEIRPORT.

unqualified

Publisher

Oxford Archaeological Unit

No need for qualification

This can be auto generated if publisher is consistent (as for SMR type data). Thus it can be 'hard-wired' into the sql access point query for publisher on Z39.50 targets and something similar for OAI

DC_Relation

Related resources such as bibliographic information or related identifiers.

unqualified

Relation Identifier

123456
654321
Oxoniensia 29-30/1964-5/191

qualified

Relation.Type	Relation.Identifier
DepositorID	123456
AssociatedID	654321
BibliographicReference	Oxoniensia 29-30/1964-5/191

These are the main types used by ADS. A depositor ID becomes a relation if data physically held by ADS (see DC_Identifier as to how a new unique identifier is created). This does not apply to remote Z39.50 targets but shift from DC_Identifier to

DC_Relation necessary for OAI data. Some SMRs record related records here AssociatedID. Some also hold Bibliographic Information.

DC_Rights

Can detail rights management statement: copyright, access, etc. Often a URL to a rights statement. Seen limited use by ADS as we have front end rights statements on interfaces. Although for HEIRPORT we also added a URL which displays within results set. This can be auto generated or hard-wired (OAI or Z39.50) as described for other elements above

unqualified

Rights

href="http://ads.ahds.ac.uk/a_c.html"

No need for qualification. This can be auto-generated or hard-wired (OAI or Z39.50) as described for other elements above with the rest of stuff to make it into a live link built into the queries to generate a results set

DC_Source

Information about a resource from which the present is derived – e.g. books, archives, datasets, etc. Normally referenced by an ISBN, ISSN, URL, etc. It is here that links can be provided back to partner databases if they are online and have more detailed information locally through using local identifier

unqualified

Source Name

1 827256 62 1

0867-6542

http://

qualified

Source.Type

Source.Name

ISBN

1 827256 62 1

ISSN

0867-6542

URL

http://adatabase.org?667654

In case of URLs rest of stuff to make it into a live link built into the queries to generate a results set

DC_Subject

Keywords classifying theme or concept of a resource – use of controlled vocabularies (thesauri) where possible. Keywords can include monument types, intervention types, data types, object types, material in case of objects, etc

unqualified

Subject
Settlement
Barrow
Monastery
Coin
Excavation
Images

qualified

ADS data is qualified by Subject Type but not essential to do as above are self evident

DC_Title

Name given to a resource

unqualified

Title
Excavations at Eynsham Abbey

Not necessary to qualify

DC_Type

Resource type – nature of a resource. Effectively for us this is a collections level descriptor. ADS use UKOLN collections level descriptors which include the following

Collection = records based collections (e.g. SMR type data)
DigitalArvchive = digital archive

Which should cover most things for us

unqualified

Type
Collection

No need to qualify

Can auto generate or hardwire as noted for other elements

The above provides guidance to Arena partners for mapping their data to Dublin Core in a consistent manner across the partnership. It is likely that data is not currently held for all of the DC elements. Where perceived as useful (for instance Language becomes an important metadata element in a multilingual environment) it can often be auto-generated or hardwired into queries

Future work

XML DTDs as used by OAI exist for unqualified Dublin Core but will need creating for the qualified DC suggested above

References

DCMES. 2003. 'Dublin Core Metadata Element Schema, version 1.1: Reference Description', DCMI <http://dublincore.org/documents/2003/02/04/dces/> (downloaded 10.3.2003)

Miller, P. & Greenstein, D. (eds). 1997. *Discovering Online Resources Across the Humanities: A Practical Implementation of the Dublin Core*, UKOLN

Office of the e-Envoy. 2003. 'e-Government Metadata Standard version 2 (draft)', [http://www.eenvoy.gov.uk/oeo/oeo.nsf/sections/consultationstop/\\$file/consultationsindex.htm](http://www.eenvoy.gov.uk/oeo/oeo.nsf/sections/consultationstop/$file/consultationsindex.htm) (downloaded 10.3.2003)

UKOLN. Collections Description Focus, <http://www.ukoln.ac.uk/cd-focus/> (downloaded 11.3.2003)