The Dublin Core Application Profile for Scholarly Works (ePrints)

Metadata issues for Scottish institutional repositories
4th December 2006, National Library of Scotland, Edinburgh

Julie Allinson
Repositories Research Officer
UKOLN, University of Bath
Overview

- Background, scope and functional requirements
- Model, application profile and vocabularies
- Next steps …
Background and scope

• Overall aim
  – to offer a solution to metadata issues identified in Eprints UK project, and others (e.g. PerX project)
  – to provide a richer metadata profile for the Intute repository search project, and others?

• Scope
  – DC elements plus any additional elements necessary
  – Identifiers for the eprint and full-text(s); related resources
  – Hospitable to the use of a variety of subject access solutions
  – Additional properties to fulfil search/browse requirements
  – Bibliographic citations and references citing other works
Deliverables

- Functional Requirements Specification
- Entity-Relationship Model
- Application Profile with Cataloguing/Usage Guidelines
- Plan for Community Acceptance and Take-up
Requirements summary

- richer metadata set - consistent metadata
- support for added-value services
- unambiguous method of identifying full-text(s)
- consider version identification and most appropriate copy of a version
- open access materials
- support browse based on controlled vocabularies
- OpenURL link servers
- support citation analysis (in line with dc-citation WG recommendations)
- identification of the research funder and project code
- identification of the repository or other service making available the copy
- date available
- date of modification of a copy, to locate the latest version

the requirements demanded a more complex model …
Model : what’s that?

- The model says what things are being described
  - the set of **entities** that we want to describe
  - and the key **relationships** between those entities
- FRBR (Functional Requirements for Bibliographic Records) provides the basis for our model
  - a model for the entities that **bibliographic records** are intended to describe
  - but we’ve applied it’s model to **scholarly works**
  - and it could be applied to other **resource types**
FRBR?

- FRBR models the world using 4 key entities: Work, Expression, Manifestation and Item
  - A **work** is a distinct intellectual or artistic creation. A work is an abstract entity
  - An **expression** is the intellectual or artistic realization of a work
  - A **manifestation** is the physical embodiment of an expression of a work.
  - An **item** is a single exemplar of a manifestation. The entity defined as item is a concrete entity.
- FRBR also defines additional entities - 'Person', 'Corporate body', 'Concept', 'Object', 'Event' and 'Place'
- And the relationships between entities
The model

ScholarlyWork

Expression

Manifestation

Agent

Copy

AffiliatedInstitution

isSupervisedBy

isFundedBy

isCreatedBy

isEditedBy

isManifestedAs

isAvailableAs

isExpressedAs

isPublishedBy

0..∞
Vertical vs. horizontal relationships
FRBR for eprints

The eprint as a scholarly work

Author’s Original 1.0
Author’s Original 1.1
...
Version of Record (English)
Version of Record (French)

html
pdf

publisher’s copy
institutional repository copy

scholarly work (work)
version (expression)
format (manifestation)
copy (item)
Capturing this in Dublin Core

- The DCMI Abstract Model (DCAM) says what the descriptions look like
- it provides the notion of ‘description sets’
- i.e. groups of related ‘descriptions’
- where each ‘description’ is about an instance of one of the entities in the model
- relationships and attributes are captured as metadata properties in the application profile
From model to profile

- the model defines the entities and relationships
- each entity and its relationships are described using an agreed set of attributes / properties
- the application profile describes these properties
  - contains recommendations, cataloguing/usage guidelines and examples
  - little is mandatory, prescriptive statements are limited
  - structured according to the entities in the model
The application profile

- simple DC properties (the usual suspects … )
  - identifier, title, abstract, subject, creator, publisher, type, language, format
- qualified DC properties
  - access rights, licence, date available, bibliographic citation, references, date modified
- new properties
  - grant number, affiliated institution, status, version, copyright holder
- properties from other schemes
  - funder, supervisor, editor (MARC relators)
  - name, family name, given name, workplace homepage, mailbox, homepage (FOAF)
- clearer use of existing relationships
  - has version, is part of
- new relationship properties
  - has adaptation, has translation, is expressed as, is manifested as, is available as
- vocabularies
  - access rights, entity type, resource type and status
Example properties

**ScholarlyWork:**
- title
- subject
- abstract
- affiliated institution
- identifier

**Expression:**
- title
- date available
- status
- version number
- language
- genre / type
- copyright holder
- bibliographic citation
- identifier

**Manifestation:**
- format
- date modified

**Agent:**
- name
- type of agent
- date of birth
- mailbox
- homepage
- identifier

**Copy:**
- date available
- access rights
- licence
- identifier
Thoughts on the approach …

• this approach is guided by the functional requirements identified and the primary use case of richer, more functional, metadata
• it also makes it easier to rationalise ‘traditional’ and ‘modern’ citations
  – traditional citations tend to be made between eprint ‘expressions’
  – hypertext links tend to be made between eprint ‘copies’ (or ‘items’ in FRBR terms)
• a complex underlying model may be manifest in relatively simple metadata and/or end-user interfaces
• existing eprint systems may well capture this level of detail currently – but use of simple DC stops them exposing it to others!
OAI-PMH

- Dumb-down
  - we still need to be able to create simple DC descriptions
  - we have chosen to dumb-down to separate simple DC descriptions of the ScholarlyWork and each Copy
    - simple DC about the ScholarlyWork corresponds to previous guidance
    - simple DC about each Copy useful for getting to full-text, e.g. by Google

- XML schema
  - http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_DC_XML
  - almost finished
Next steps …

• Community acceptance plan outlines further work towards community take-up
  – deployment by developers
  – deployment by repositories, services
  – dissemination
  – possible establishment of a DC task group

• More application profiles
  – JISC is funding work on profiles for images, time-based media and geographic data
  – this approach may prove a good foundation
Thanks …

www.ukoln.ac.uk/repositories/
digirep/index/Eprints_Application_Profile

<j.allinson@ukoln.ac.uk>
<andy.powell@eduserv.org.uk>

Presentation available from the above URL