

# **The Eprints Application Profile: a FRBR approach to modelling repository metadata**

## **OR 2007**

the 2<sup>nd</sup> International Conference on Open Repositories

**San Antonio, Texas, USA**

**23-26 January 2007**

Julie Allinson, UKOLN, University of Bath

Pete Johnston, Eduserv Foundation

Andy Powell, Eduserv Foundation

# abstract

Julie Allinson, Pete Johnston and Andy Powell, UKOLN, University of Bath, present recent work on developing a Dublin Core Application Profile (DCAP) for describing "scholarly publications" (eprints). They will explain why **the Dublin Core Abstract Model is well suited to creating descriptions based on entity-relational models** such as the FRBR-based (Functional Requirements for Bibliographic Records) Eprints data model. The ePrints DCAP highlights the relational nature of the model underpinning Dublin Core and illustrates that **the Dublin Core Abstract Model can support the representation of complex data describing multiple entities and their relationships.**

# overview

- background, scope and functional requirements
- the model
- the application profile and vocabularies
- oai-pmh , dumb-down and community acceptance

# background, scope and functional requirements

# terminology

- eprints, research papers and scholarly works are used synonymously for
  - a "scientific or scholarly research text"  
(as defined by the Budapest Open Access Initiative  
<http://www.earlham.edu/~peters/fos/boaifaq.htm#>  
)
  - e.g. a peer-reviewed journal article, a preprint, a working paper, a thesis, a book chapter, a report, etc.
- the application profile is independent of any particular software application

# the problem space

- simple DC is insufficient to adequately describe eprints
- the metadata produced is often inconsistent and poor quality
- identifying the full-text is problematic
- this poses problems for aggregator services

# the work

- the work aimed to develop:
  - a Dublin Core application profile for eprints ;
  - any implementation / cataloguing rules to support functionality offered by the Intute repository search service, such as fielded searches of the metadata or indexing the full-text of the research paper;
  - a plan for early community acceptance and take-up, bearing in mind current practice
- co-ordinated by Julie Allinson (UKOLN) and Andy Powell (Eduserv Foundation), summer 2007
- through a working group and feedback group
- using a wiki to make all documentation freely available, at all times

## the scope

- as provided by JISC, the funders
  - DC elements plus any additional elements necessary
  - identifiers for the eprint and full-text(s), and related resources
  - hospitable to a variety of subject access solutions
  - additional elements required as search entry points
  - bibliographic citations and references citing other works



# the functional requirements : a selection

- richer metadata set & consistent metadata
- unambiguous method of identifying full-text(s)
- version identification & most appropriate copy of a version
- identification of open access materials
- support browse based on controlled vocabularies
- OpenURL & citation analysis
- identification of the research funder and project code
- identification of the repository or 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 ...**

# the model

## what is an application model?

- the application model says what things are being described
  - the set of **entities** that we want to describe
  - and the key **relationships** between those entities
- **model vs. Model** - the application model and the DCMi Abstract Model are completely separate
- the DCMi Abstract Model says what the descriptions look like (more later ... )

# FRBR

- FRBR (Functional Requirements for Bibliographic Records) is a model for the entities that ***bibliographic records*** are intended to describe
- 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

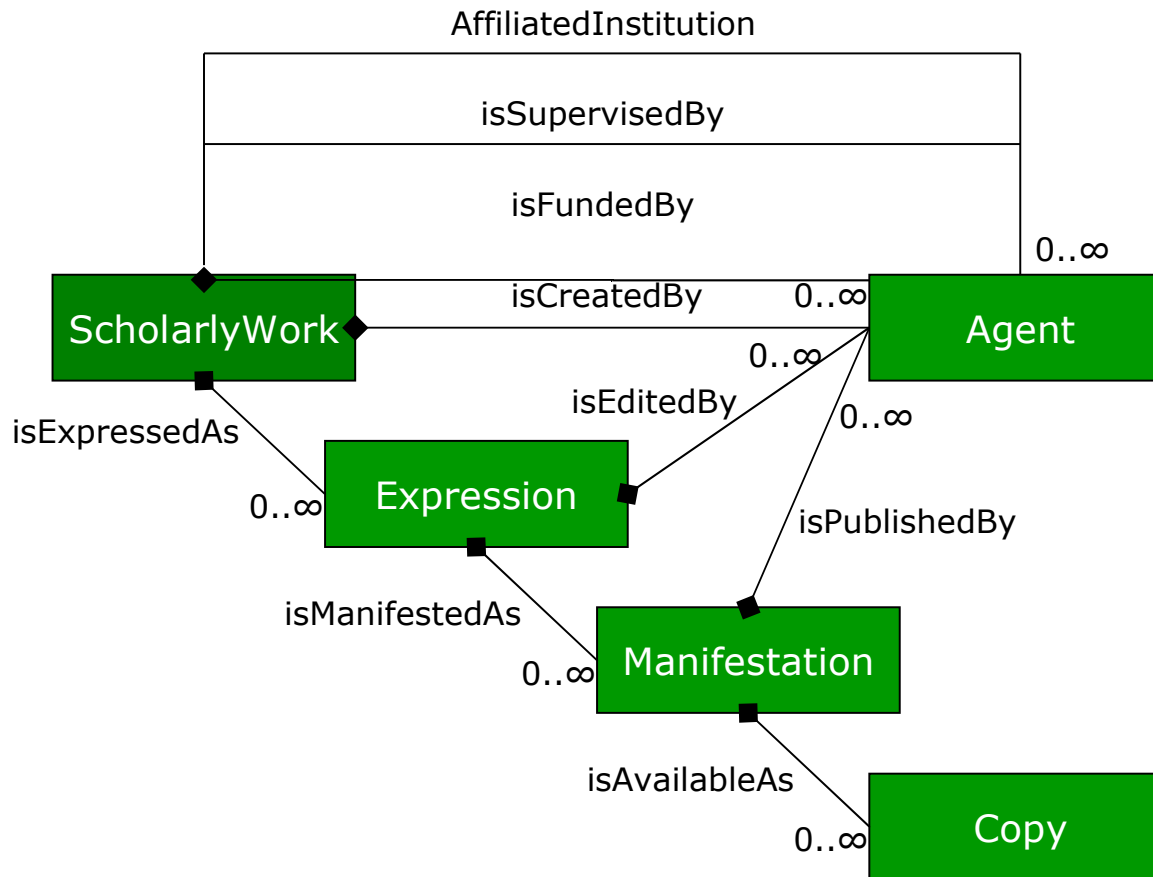
# FRBR relationships

- FRBR also defines additional entities that are related to the four entities above - 'Person', 'Corporate body', 'Concept', 'Object', 'Event' and 'Place' - and relationships between them
- the key entity-relations appear to be:
  - Work -- is realized through --> Expression
  - Expression -- is embodied in --> Manifestation
  - Manifestation -- is exemplified by --> Item
  - Work -- is created by --> Person or Corporate Body
  - Manifestation -- is produced by --> Person or Corporate Body
  - Expression -- has a translation --> Expression
  - Expression -- has a revision --> Expression
  - Manifestation -- has an alternative --> Manifestation

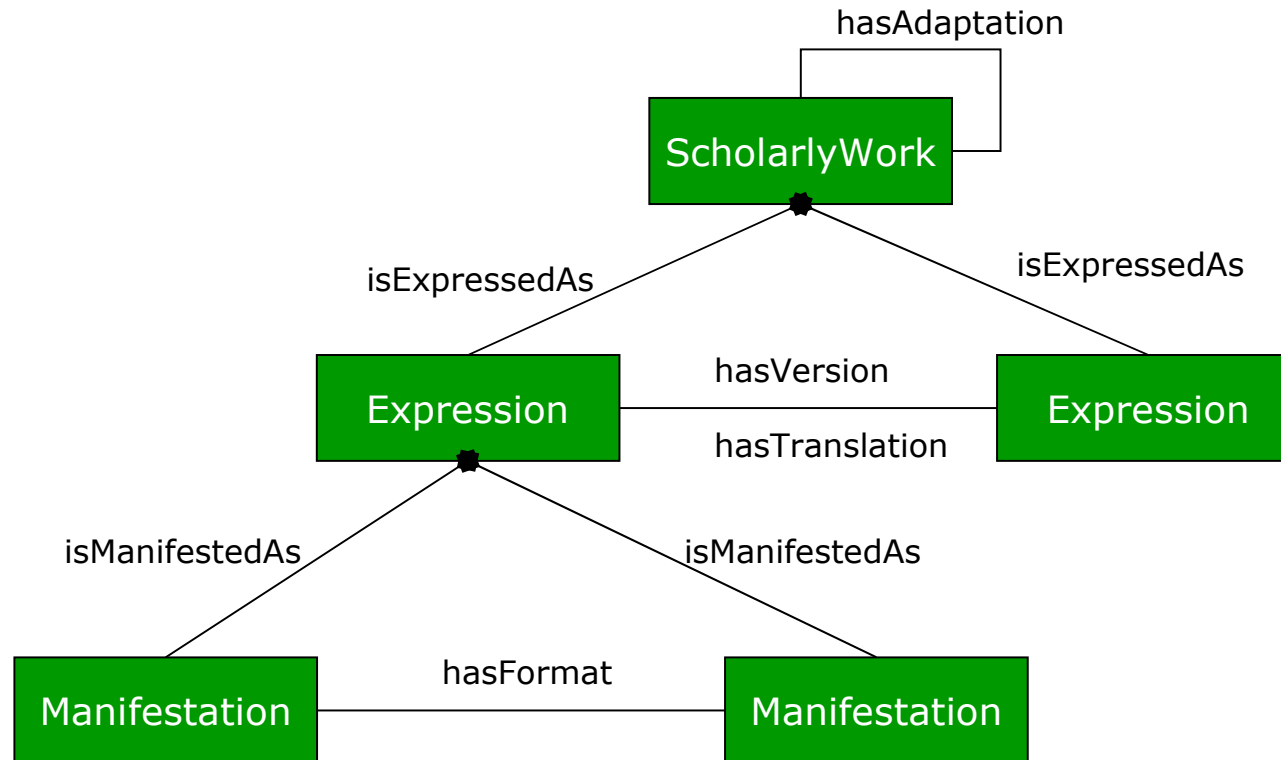
# FRBR for eprints

- FRBR provides the basis for our model
  - it's a model for the entities that ***bibliographic records*** describe
  - but we've applied it to ***scholarly works***
  - and it might be applied to other ***resource types***
- FRBR is a useful model for eprints because it allows us to answer questions like:
  - what is the URL of the most appropriate copy (a FRBR item) of the PDF format (a manifestation) of the pre-print version (a expression) for this eprint (the work)?
  - are these two copies related? if so, how?

# the model



# vertical vs. horizontal relationships







# the paper : multiple expressions, manifestations and copies

scholarly work  
(work)

version  
(expression)

format  
(manifestation)

copy  
(item)

Signed metadata paper  
(the eprint as scholarly work)

Author's  
Original 1.0

Author's  
Original 1.1

...

Version of  
Record  
(English)

Version of  
Record  
(Spanish)

pdf

doc

print copy

pdf

html

institutional  
repository  
copy

published  
proceedings

publisher's  
repository  
copy

institutional  
repository  
copy

author's  
web site  
copy

no digital copy available  
(metadata only)

## the presentation : expression(s) or new scholarlyWork?

Slides  
(ppt)

### Signed metadata : method and application

International Conference on Dublin Core and Metadata Applications, 3 – 6 October 2006, Mexico

Julie Allinson (presenter)  
Repositories Research Officer  
UKOLN, University of Bath

Emma Tonkin (author)  
Interoperability Focus Officer  
UKOLN, University of Bath



UKOLN is supported by: JISC M L A  
Museums, Libraries and Archives Council



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

A centre of expertise in digital information management

flickr

Home Learn More Sign Up! Explore

### Signed-metadata-dublin-core

Created by [bostonjail](#).

Flickr  
(jpeg)

Conclusion

- of provenance and identity are dealt with in the present digital library realm by the perceived integrity of a source
- As the number of metadata sources and aggregators increase, these informal mechanisms may prove insufficient and metadata may be subject to abuse
- Digitally signing metadata records can help to identify provenance
- Public key infrastructure functionality offers particular cryptographic methods to digitally signing metadata
- And can help to create new networks of trust



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)  
A centre of expertise in digital information management

17 photos | [Detail view](#)

Photos are from  
17 Nov 06...



slideshare BETA

Home My Slidespace Upload To  
Latest Popular Featured Most Commented

Power  
London  
Creatio  
Ads by

Slideshare  
(what format?)

### Signed metadata : met

Repositories Research Officer  
UKOLN, University of Bath

Emma Tonkin (author)  
Interoperability Focus Officer  
UKOLN, University of Bath



[www.ukoln.ac.uk](http://www.ukoln.ac.uk)

UKOLN A centre of expertise in digital information management

UKOLN is supported by: JISC M L A  
Museums, Libraries and Archives Council



[www.bath.ac.uk](http://www.bath.ac.uk)

Tags

No tags to display



audio

## capturing this in DC

- 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 application 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

# application profile and vocabularies

# The application profile

- DC Metadata Element Set properties (the usual simple DC suspects ... )
  - identifier, title, abstract, subject, creator, publisher, type, language, format
- DC Terms properties (qualified DC)
  - access rights, licence, date available, bibliographic citation, references, date modified
- new properties
  - grant number, affiliated institution, status, version, copyright holder
- properties from other metadata property sets
  - 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



# oai-pmh , dumb-down and community acceptance

# OAI-PMH, dumb-down

- 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 is useful for getting to full-text, e.g. by Google
- XML schema
  - produced by Pete Johnston, Eduserv Foundation
  - specifies an XML format (Eprints-DC-XML) for representing a DC metadata *description set*
  - based closely on a working draft of the DCMI Architecture Working Group for an XML format for representing DC metadata (DCXMLFULL)
  - enables the creation, exposure and sharing of Eprints DC XML (epdcx)

# community acceptance

- community acceptance plan outlines further work towards community take-up
  - deployment by developers
  - deployment by repositories, services
  - dissemination
  - DC task group may take forward development of the profile
- more application profiles
  - JISC is funding work on profiles for images, time-based media and geographic data
  - this approach may prove a good foundation

## 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!
- it is the DCAM that allows us to do this with Dublin Core

# Thank you

- Profile, presentation and all documentation available from:

**<http://www.ukoln.ac.uk/repositories/digirep/>**

- Julie Allinson [j.allinson@ukoln.ac.uk](mailto:j.allinson@ukoln.ac.uk)
- Pete Johnston [pete.johnston@eduserv.org.uk](mailto:pete.johnston@eduserv.org.uk)
- Andy Powell [andy.powell@eduserv.org.uk](mailto:andy.powell@eduserv.org.uk)