

JISC CETIS Metadata and Digital
Repository SIG meeting,
Manchester

16 April 2007

A Dublin Core Application Profile for Scholarly Works (eprints)

Julie Allinson

Repositories Research Officer
UKOLN, University of Bath



Supported by

JISC



Museums, Libraries and
Archives Council

www.ukoln.ac.uk

A centre of expertise in digital information management



UNIVERSITY OF
BATH



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
www.earlham.edu/~peters/fos/boaifaq.htm#liter
)
 - 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
 - containing properties 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;
 - any implementation / cataloguing rules;
 - 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 2006
- 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

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

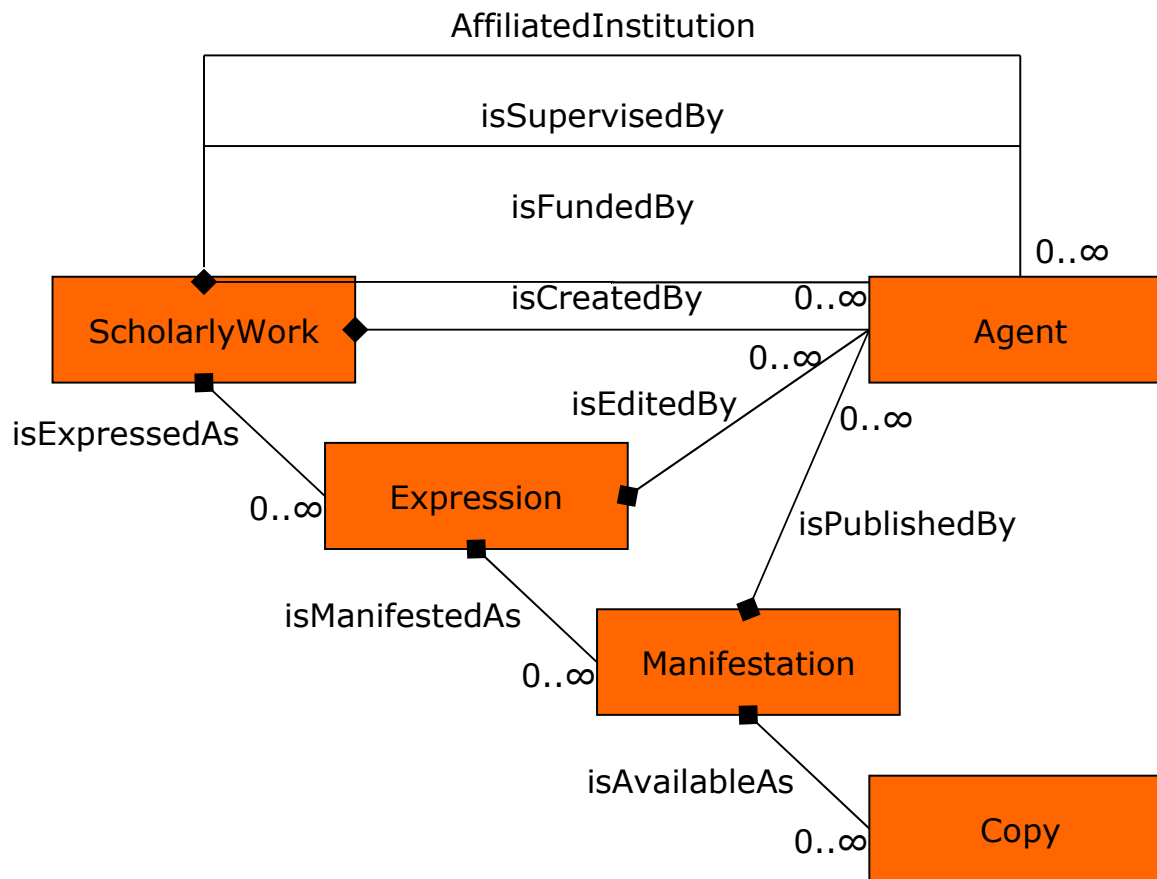
RBR relationships

- 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

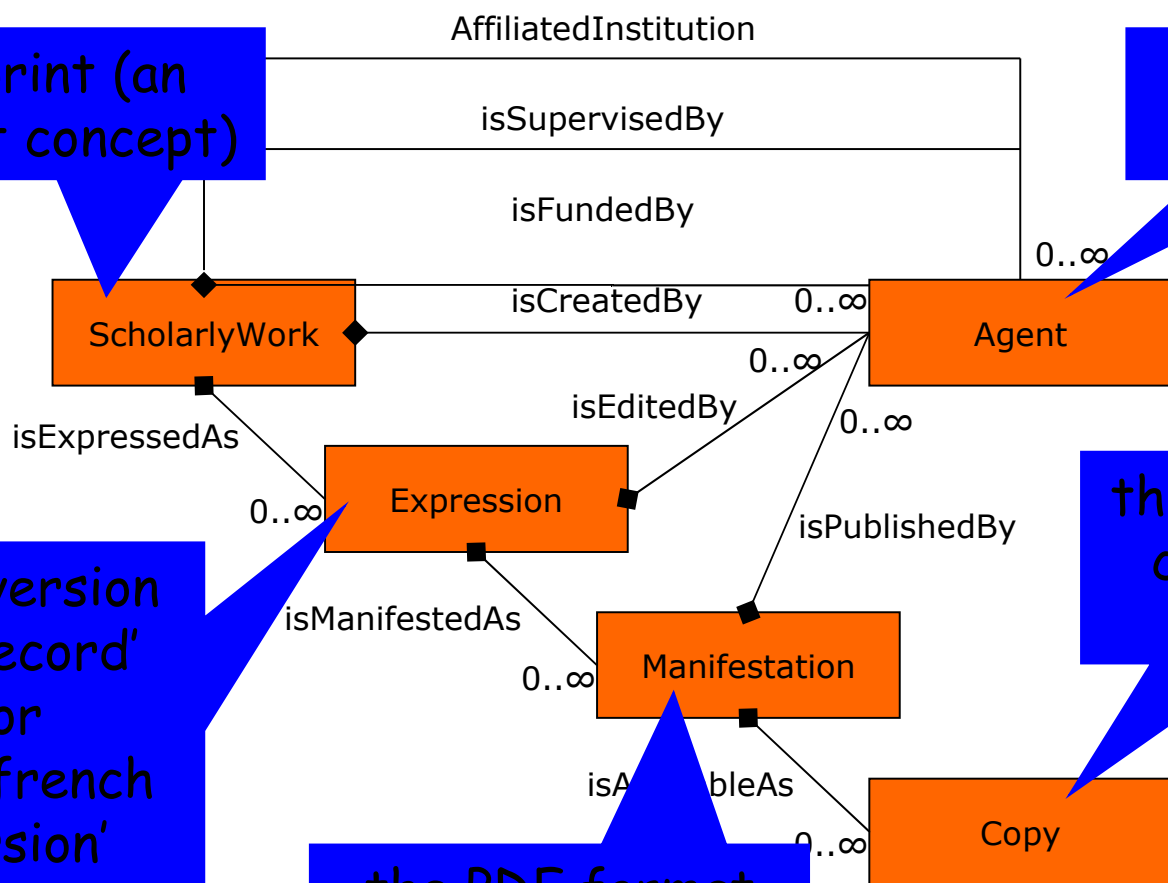
FRBR for eprints

- 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



the model



the eprint (an abstract concept)

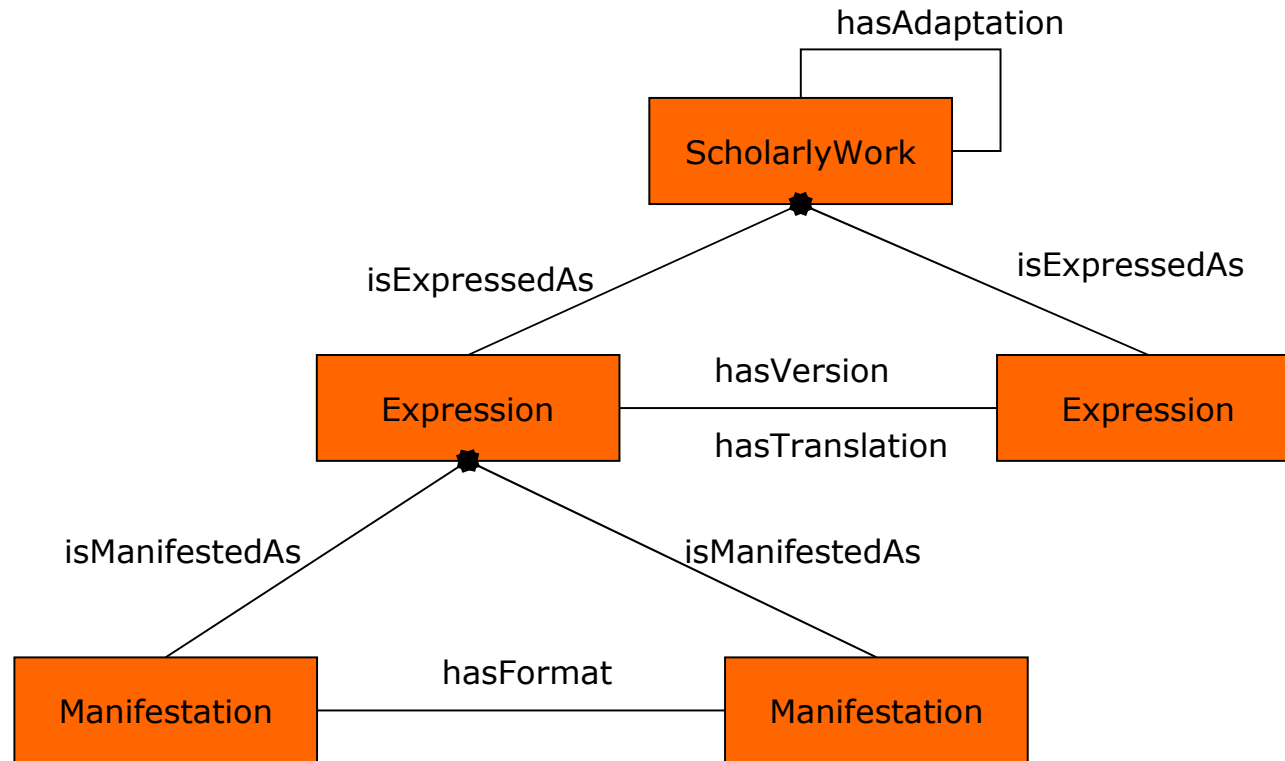
the author or the publisher

the 'version of record' or the 'french version' or 'version 2.1'

the PDF format of the version of record

the publisher's copy of the PDF ...

vertical vs. horizontal relationships



an example - a conference paper

Signed metadata: method and application

Emma Tonkin¹

Abstract

As metadata providers increase in number and diversity, and additional contexts for metadata use are identified, issues of trust, provenance and identity gain in relevance. Use of a Public-Key Infrastructure (PKI) is discussed for digital signature of metadata records, providing evidence of the identity of the signer and the authenticity of the information within the record. Two methods are suggested; firstly, the W3C XML Signature, and secondly, identification of a minimal set of metadata elements that enable signature verification across various character sets and formats, using the OpenPGP standard. Possible strategies for handling annotation within this infrastructure are suggested. Finally, some use cases are briefly discussed.

Keywords

heterogeneous infrastructure, digital signature, web of trust, provenance.

1. Introduction

The issue of trust, a level of confidence in a source, is of great importance on the Internet in general. The source of a piece of information is a vital detail in analysis; is the source known? Do they generally provide accurate information? Do they have a reason to provide inaccurate information? In this manner, the provenance of a piece of information becomes a necessary detail in analysis and interpretation.

The predominance of the client-server model means that this issue may often be ignored either partially or wholly, particularly in the digital library environment, in that metadata providers are considered to be responsible for the accuracy of their content. Provenance is established either implicitly, or explicitly stated within metadata; the Open Archives Initiative provides the <provenance> tag, permitting versioning of metadata across systems. The model has been further refined in various contexts, such as the DART project (Dahlquist *et al.*), where the model relies on the accuracy of the metadata provider's data; in-
the number of intermediate organisations
relatively low, composed mostly
the responsibilities

This paper discusses the role that PKI digital signatures can play in provenance and accountability to be handled.

1.1 Principles behind the digital signature

The digital signature, first introduced in Diffie and Hellman (1976), is a cryptographic technique, that aims to permit the verification of a message as received is equal to the message as sent. A digital signature also requires as a prerequisite that the sender of the message verify that a handwritten signature has not been used. A similar prerequisite exists for the use of the digital signature as an external reference to known identities, and to demonstrate

This is possible using a variety of cryptographic techniques, today is based around PKI, in which the process of creating the digital signature, is known by those who need the ability to verify the signature. A public key cannot be used to sign a message, as any member of the public to verify the signature the private key remains secret.

Public keys are often distributed and corresponding identity information is required. It is of course possible to distribute identity information, in an analogy is not a general solution to a known identity is established by creating a signature as being the identity has signed the message. However, if a key is compromised and the message originates from the work of trust.

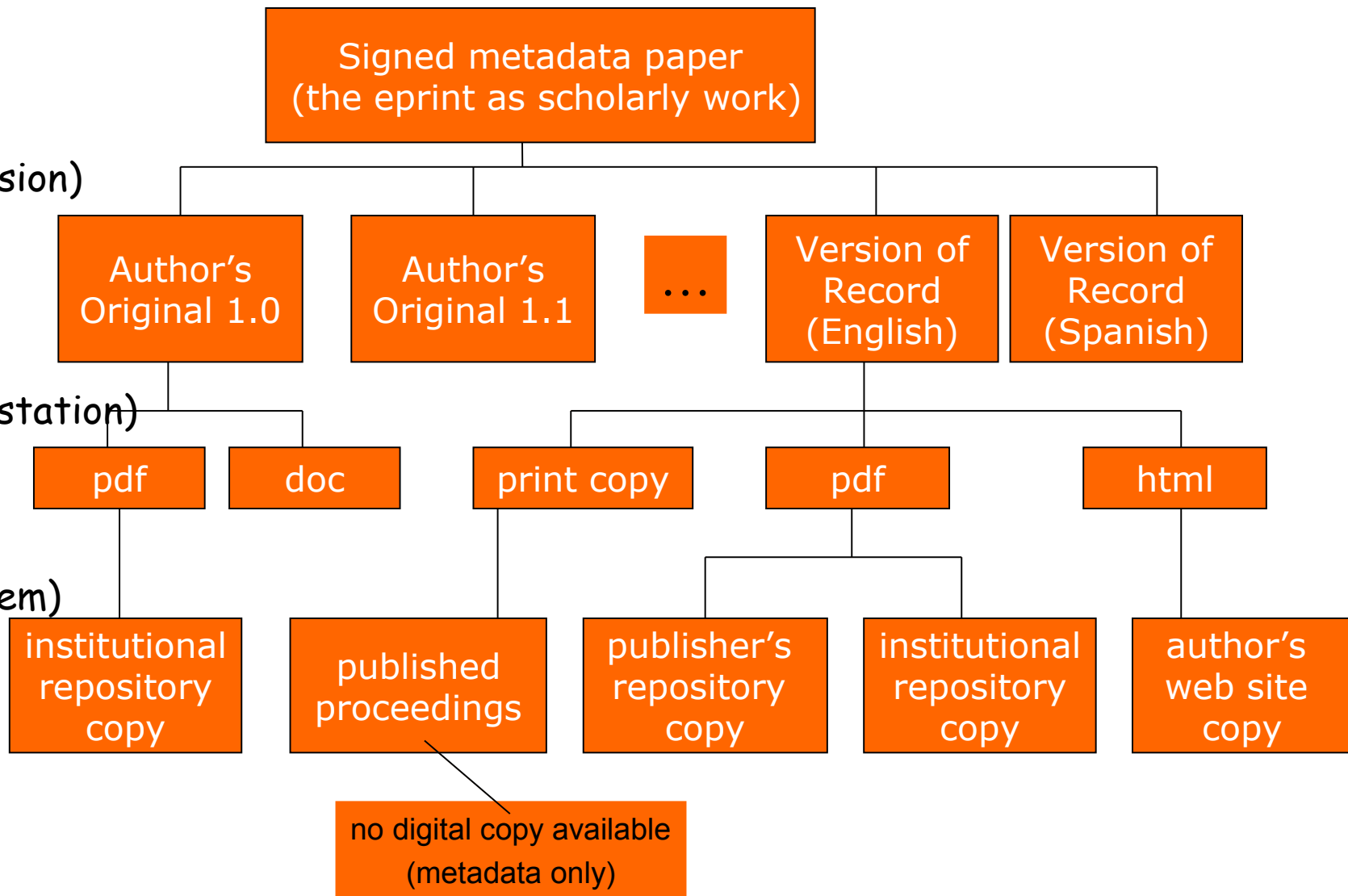
the paper : multiple expressions, manifestations and copies

scholarly work
(work)

version
(expression)

format
(manifestation)

copy (item)



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** **MLA**
Museums, Libraries and Archives Council
UNIVERSITY OF BATH
www.bath.ac.uk

Slideshare
(what format?)



Signed metadata : met

Repositories Research Officer
UKOLN, University of Bath

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

UKOLN is supported by: **JISC** **MLA**
UNIVERSITY OF BATH
www.bath.ac.uk

UKOLN A centre of expertise in digital information management

share 1 / 17

flickr

Home Learn More Sign Up! Explore

Signed-metadata-dublin-core

Created by bostonjal.

Flickr
(jpeg)

Conclusion

- of provenance and identity are dealt with in the digital 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
A centre of expertise in digital information management

17 photos | [Detail view](#)

Photos are from
17 Nov 06...

Tags

No tags to display

audio

capturing this in DC

- the DCM I 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 community *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!

Julie Allinson j.allinson@ukoln.ac.uk
www.ukoln.ac.uk/repositories/digirep/