Repositories and the wider context

Exchange of Experience on Institutional/Digital Repositories

3 November 2006, Liverpool

Julie Allinson

Repositories Research Officer UKOLN, University of Bath

UKOLN is supported by:



Supported by



Museums, Libraries and Archives Council



Overview

- JISC Information Environment / repository ecology (a little)
- Supporting repository infrastructure (a little)
- JISC funding for repositories (a little)
- Dublin Core Application Profile for Scholarly Works (more)
 - background, scope and functional requirements
 - model, application profile and vocabularies
 - next steps …



What is the JISC Information Environment?

- Development activity
- Funding stream
- Technical architecture
 - relationships between components, standards, protocols
- Content and data flow
- User's experience of networked information in educational context



JISC Information Environment

"The aim of the Information Environment is to help provide convenient access to resources for research and learning through the use of resource discovery and resource management tools and the development of better services and practice."

JISC Draft Strategy 2007-2009





This work is licensed under a <u>Creative Commons License</u> Attribution-ShareAlike 2.0



Repository ecology

- Increasingly complex repository ecology
 - interactions between repositories ...
 - between repositories and services ...
 - between repositories and users
 - UKOLN are mapping this area
- Forthcoming repository interfaces working group facilitating interoperability between a distributed network of repositories
- Interoperability is more than harvesting metadata
 - Data exchange, tracking provenance, added value services, enabling re-use
- Feeding into Object Reuse and Exchange (ORE) initiative www.openarchives.org/ore



Supporting repository infrastructure

- Three recently JISC-funded projects:
 - Repository Support Project
 - Sherpa (lead), UKOLN, University of Aberystwyth, University of Southampton
 - Aim: to provide support for institutions setting up repositories
 - Intute Repository Search Project
 - Intute (lead), UKOLN, Sherpa
 - Aim: to provide a cross search of UK repositories
 - www.intute.ac.uk/projects.html
 - PROSPERO / The Depot ("put it in the depot")
 - Edina (lead), Sherpa
 - Aim: To point depositors to an appropriate repository
 - ... and to provide an interim repository for depositors without an institutional or other repository
 - www.edina.ac.uk/projects/prospero/



JISC funding for repositories

- Digital Repositories Programme ('05-'07)
 - 25 projects
 - less focus on 'institutional' repositories (than FAIR Programme)
 - some themes: legal/policy, culture/social, research data, e-learning, technical challenges, interoperability
- Capital Programme: Repositories and Preservation Strand ('06-'09)
 - includes the three projects already mentioned ... and 11 more
 - current funding circular
 - Tools and innovation
 - Discover to delivery
 - Repository start-up and enhancement
 - Digital preservation and records management
 - Shared infrastructure services
 - deadline 23rd November



Repositories Research Team

- Working across both Programmes
- UKOLN, CETIS collaboration
- Cross-programme synthesis and support
- Research activities
- Communication and dissemination (like today)
- Working closely with JISC Programme Managers, but we aren't the JISC!



Dublin Core Application Profile for Scholarly Works (ePrints)



Background and scope

- Not just for eprints.org software ... platform independent
- Overall aim
 - to offer a solution to metadata issues identified in Eprints UK project, and by others (e.g. PerX project)
 - to provide a richer metadata profile for the Intute repository search service
- Scope
 - Dublin Core elements plus any additional elements necessary
 - Identifiers for the eprint and full-text(s); related resources etc
 - Hospitable to the use of a variety of subject access solutions
 - Bibliographic citations and references citing other works

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



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
 - an 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





Vertical and horizontal relationships



FRBR for eprints



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, affiliation 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				Agent:	
ScholarlyWork: title				name type of agent	
subject abstract affiliated institu	Expression: title date available			mailbox homepage identifier	
Identiner	status version number language	Ma forr	Manifestation: format		
copyright holde bibliographic cit identifier		dat tion	e modified	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!



Next steps ...

- Dumb-down
 - we still need to be able to create simple DC descriptions for OAI-PMH
 - we have chosen to dumb-down to 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
- Community acceptance plan outlines further work towards community take-up
 - xml schema (awaiting new Dublin Core XML guidelines)
 - deployment by developers
 - deployment by repositories, services
 - Dissemination
- More application profiles
 - JISC is funding work on profiles for images, time-based media and geographic data





And that's it ...

Find out more:

- www.ukoln.ac.uk/repositories
- www.jisc.ac.uk/whatwedo/themes/information_environment

Contact:

• j.allinson@ukoln.ac.uk

This presentation soon to be available at:

www.ukoln.ac.uk/repositories/digirep/index/Media:nowal-repositories-exchange.ppt

