An introduction to …

Repository reference models

CETIS Metadata and Digital Repositories SIG Meeting
1st March 2006, HE Academy, York

Julie Allinson
Digital Repositories Support Officer
UKOLN, University of Bath

UKOLN is supported by:

JISC

Supported by
MLA
Museums, Libraries and Archives Council

www.ukoln.ac.uk

UNIVERSITY OF BATH

www.bath.ac.uk

Overview

- Reference models, context and background
- JISC Digital Repositories Programme, scoping a repository reference model and other activities
- “Repositories …. and the known unknowns”
  <www.e-framework.org/events/conference/programme/repositories/>

Reference models

context and background

“Reference models, why bother?”

- Bill Olivier

- "A "Reference Model" **bridges** :
  - the World of Users and their Work
  - the underlying Technical Services and their associated Specs (The Wall)"

Bill Olivier, 'Why "Reference Models"?, JISC CETIS Conference 2005

<www.e-framework.org/events/conference/programme/Why_Ref_Models.ppt>

A reference model includes …

- A human domain map / model / ontology
  - The stakeholders and their purposes
  - The functions and tasks
  - The entities used and their relationships
- The patterns of practices, processes and workflows
  - As is (identifying good practices & problems areas)
  - To be (innovations &/or improvements)
- The technical part
  - The user interface, tool or application layer
  - The services that lie behind
  - A co-ordination layer
    - An “orchestration” of services
    - A service “choreography”

Levels of abstraction

- High level [informative]
  - Domain maps, models, ontologies
  - OAIS is a high-level reference model
- Medium level
  - Processes, practices, workflows, scenarios, use cases, context and systems
- Low level [normative]
  - Specifications (e.g. SCORM), orchestration / choreographic models, reference implementations, architectures

Role within the e-Framework

- a service-oriented description of a collection of processes and workflows supporting at least one application
- may cover one or more domains
- may be related to one or more other reference models
- may be incorporated into other reference models
- may overlap with the domains of one or more other reference models

- **“a core element of the e-Framework”**

JISC Digital Repositories Programme

scoping a repository reference model and other activities

Digital Repositories Programme

- 25 Projects
- Cross-domain
  - Research, e-learning, management, administration, information services …
- Range of issues
  - Technical, social, business, cultural, policy …
- Project clusters

Digital Repositories Support

- Team
  - Rachel Heery, Mahendra Mahey, Julie Allinson (UKOLN)
  - Sarah Currier, Mark Power (CETIS)
- Activities, include:
  - Process-oriented outputs : gathering, sharing, standardising
    - Scenarios, use cases, workflows
  - Repositories reference model : scoping
  - Repository ecology / typology
  - Synthesis, dissemination and communication
    - DigiRep (support wiki): www.ukoln.ac.uk/repositories/digirep/
    - JISCMAIL discussion list: JISC-REPOSITORIES@jiscmail.ac.uk

“Repositories …. and the known unknowns”

JISC CETIS Conference 2005

Key tasks

- Construct typology / ecology of repositories.
- Emerging from that typology identify common repository services and distinguish these from domain specific services.
- Identify what kind of services these repositories will offer and consume.

Current landscape

- Many existing software platforms for repositories…
- …with widespread deployment.
- Not developing software & systems from scratch.
- eFramework needs to relate to current practice.
- Reference models must accommodate existing systems.

Repositories are everywhere

- Repository issues are cross-domain and cross-community, e.g.
  - Item banks in assessment
  - ePortfolios as repository
  - document management for course validation
  - ...
- Repository reference models will overlap with other reference models

Some issues

- End users are joining up networked and desktop services to suit their own requirements.
- Personalisation is becoming a reality.
- Services may be activated at multiple points in a workflow.
- How does this relate to repository typology / ecology?

DRP projects are looking at these issues, and more.

Representation of reference models

- Use of UML for gathering requirements use cases is questionable.
- Not trying to build monolithic software applications.
- Aim is not to develop software but identify services.
- If reference models are communication tools other forms of representation more appropriate e.g. mind maps.

Repository : a definition

“ … a digital repository is differentiated from other digital collections by the following characteristics:

- content is deposited in a repository, whether by the content creator, owner or third party
- the repository architecture manages content as well as metadata
- the repository offers a minimum set of basic services e.g. put, get, search, access control
- the repository must be sustainable and trusted, well-supported and well-managed”

Heery, Rachel and Anderson, Sheila. Digital Repositories Review, UKOLN and AHDS, 2005 (Final version)
<
www.jisc.ac.uk/uploaded_documents/digital-repositories-review-200
>

Some consensus…

- Inclusive definition of “repository”.
- A little dissent…
  - Is a database a repository?
  - Managed and trusted part of this definition.
- Is such inclusive definition of “repository” useful?

A comparison of repository types

- A national LO repository, e.g. JORUM.
- An assessment item bank.
- A community image store, e.g. Flickr.
- May use similar abstract services
  - e.g. deposit, license, annotate, discover, authentication, harvest …
- But the way these services are instantiated varies enormously…
- As do the rules and policies associated with these repositories.

Rules and policies

- How do rules and policies relate to reference models?
- In the way they influence the instantiation of abstract services.
- Each rule or policy e.g. student access rights, must have one of more abstract service associated with it, e.g. authentication, authorisation.

Repository reference model(s) — why bother

- Communication tool
  - between domains
  - new developers, repository implementers
- Evolve to reflect practice, not necessarily to drive it.
- Gap analysis

The danger…

- Don’t need to retrofit a reference model to what we already know.
- Is it constructive to focus purely on the abstract?
- Focusing too much on reference models may distract us from real problems that need to be solved.
- We may just reinvent the OAIS model.

Existing reference models

- OAIS
  <www.ukoln.ac.uk/repositories/digirep/index/OAIS>
- JISC IE Discovery to Delivery (D2D) Reference Model
  <www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/dlf/>
  - Draft for discussion, by Andy Powell, based on
- DLF Service Framework
  <www.diglib.org/architectures/serviceframe/>
  - Identifies **abstract services** to **business functions** that fulfil a **business requirement**
- JISC reference models projects
  - COVARM, FREMA, eP4LL, LADIE, SCRI

  <www.elearning.ac.uk/del/refmodels>
- More …
  - see Dan Rehak's presentation at the JISC-CETIS Conference for further examples

    <www.e-framework.org/events/conference/programme>

The unknown…

As we know,
There are known knowns.
There are things we know we know.
We also know
There are known unknowns.
That is to say
We know there are some things
We do not know.
But there are also unknown unknowns,
The ones we don't know
We don't know.

The Rumsfeld approach to reference models

The solution?

- Use OAIS as our high level repository reference model.
- Use this as a communication tool across domains.
- And to help identify problem areas - the known unknowns.

The known unknowns

For example:

- Deposit API
- handling complex objects
- packaging
- federation
- identifiers
- integration with other systems

Deposit - a known unknown (1)

- We don’t know what kind of API we need to deposit into repositories.
- Flickr and Fedora have published APIs that anyone can write to.
- Can not do the same thing for Dspace or ePrints for example.

Deposit - a known unknown (2)

- Known specification relevant to deposit service binding
  - WebDAV,
  - OKI OSIDs,
  - JSR 170 & 283,
  - SRW Update,
  - Flickr API, Fedora API, ECL,...

Way forward : Deposit API

- Deposit API working group comprising a small number of developers (ePrints.org, Dspace, Fedora, ARNO, Intrallect, HarvestRoad, OCLC) to agree a deposit API.
- <www.ukoln.ac.uk/repositories/digirep/index/Deposit_API>

Way forward : OAIS

- Evaluate OAIS as a reference model for JISC repositories community.
  - Digital Repositories Programme wiki
    <www.ukoln.ac.uk/repositories/digirep/index/OAIS>
  - CETIS Metadata and Digital Repositories SIG meeting
    <metadata.cetis.ac.uk/sig_meetings/>

Unknown unknowns

- ?

http://www.ukoln.ac.uk/repositories/digirep/index/
Media:Repositories_ref_models_MDRSIG.ppt