The ARROW Project at 3 years: Looking Backwards, Aiming Forwards

David Groenewegen,
Project Manager

Andrew Treloar,
Technical Architect

The ARROW Project is funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education.

arrow.edu.au

The ARROW Consortium comprises Monash University [lead institution], National Library of Australia, The University of New South Wales and Swinburne University of Technology.

arrow
australian research repositories online to the world

MONASH University | NATIONAL LIBRARY OF AUSTRALIA | UNSW | SWINBURNE UNIVERSITY OF TECHNOLOGY

The ARROW Project

- Funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education.
- Funded 2004-2006, extended to end of 2007
- Initial objectives:

"The ARROW project will identify and test software or solutions to support best practice institutional digital repositories comprising e-prints, digital theses and electronic publishing."

ARROW Functionality Overview

- Institutional repository handling a range of inputs
  - broadly comparable in functionality with Dspace/Eprints
  - aiming to move from document objects to datasets and multimedia as well
- National research discovery service harvesting metadata from a range of repositories
  - ARROW Discovery Service
  - http://search.arrow.edu.au/

Repository Decision

- After careful analysis of available contenders in 2003 decided to go with Fedora because it has
  - robust, well architected underlying platform
  - flexible object-oriented data model
  - persistent identifiers down to the level of individual datastreams, accommodating ARROWs' compound content model
  - ability to version both content and disseminators (think of software behaviours for content)
  - clean and open exposure of APIs with well-documented SOAP/REST web services.
  - http://andrew.treloar.net/research/publications/ausweb04/
- The F in Fedora is flexible, and so...
- This choice drove need for a number of other decisions (reviewed in this presentation)

Repository Retrospective

- Flexibility is both a blessing and a curse
  - Yoga, anyone?
  - Decisions, decisions...
- Still happy to have gone with Fedora, although software delays have caused us some problems
- Fedora vision moving forward is still
  - compelling
  - well-aligned with ARROW's requirements

Metadata Decision

- Decided that Simple/Qualified DC was too limiting
  - this was prior to the DC Abstract Data Model
- Decided instead to support and store the metadata generated by communities of practice to accompany their different digital objects
- Transform MARCXML and ETD-MS metadata into Dublin Core for OAI-PMH and internal purposes
- Investigating possibility of using OCLC's interoperable core to support other transforms
- ARROW has also created some mapping transforms

Metadata Retrospective

- Still grappling with need to ensure quality metadata:
  - enforcing appropriate schema for given object type
  - managed lookups for things like names
  - controlled vocabularies (thesauri, classification schemes)
- OCLC collaboration has taken a very long time to progress
- Currently reviewing original decision in favour of
  - DC eprints profile or derivative?
  - MODS?
  - MordorMetadata™ (one scheme to tag them all...)?

Identifier Decision

- Decided to assign persistent identifiers to objects and object components (Fedora datastreams)
  - minimum persistently citeable unit can be made as granular as is required
  - repository managers can disaggregate and re-aggregate objects as required
- After careful review of alternatives decided to adopt CNRI Handles
- ARROW Handles mostly have ‘project branding’ through the resolver in the published identifier:
  - http://arrow.edu.au/hdl/1959.1/1234/
  - one site has gone instead with hdl.handle.net

Identifier Retrospective

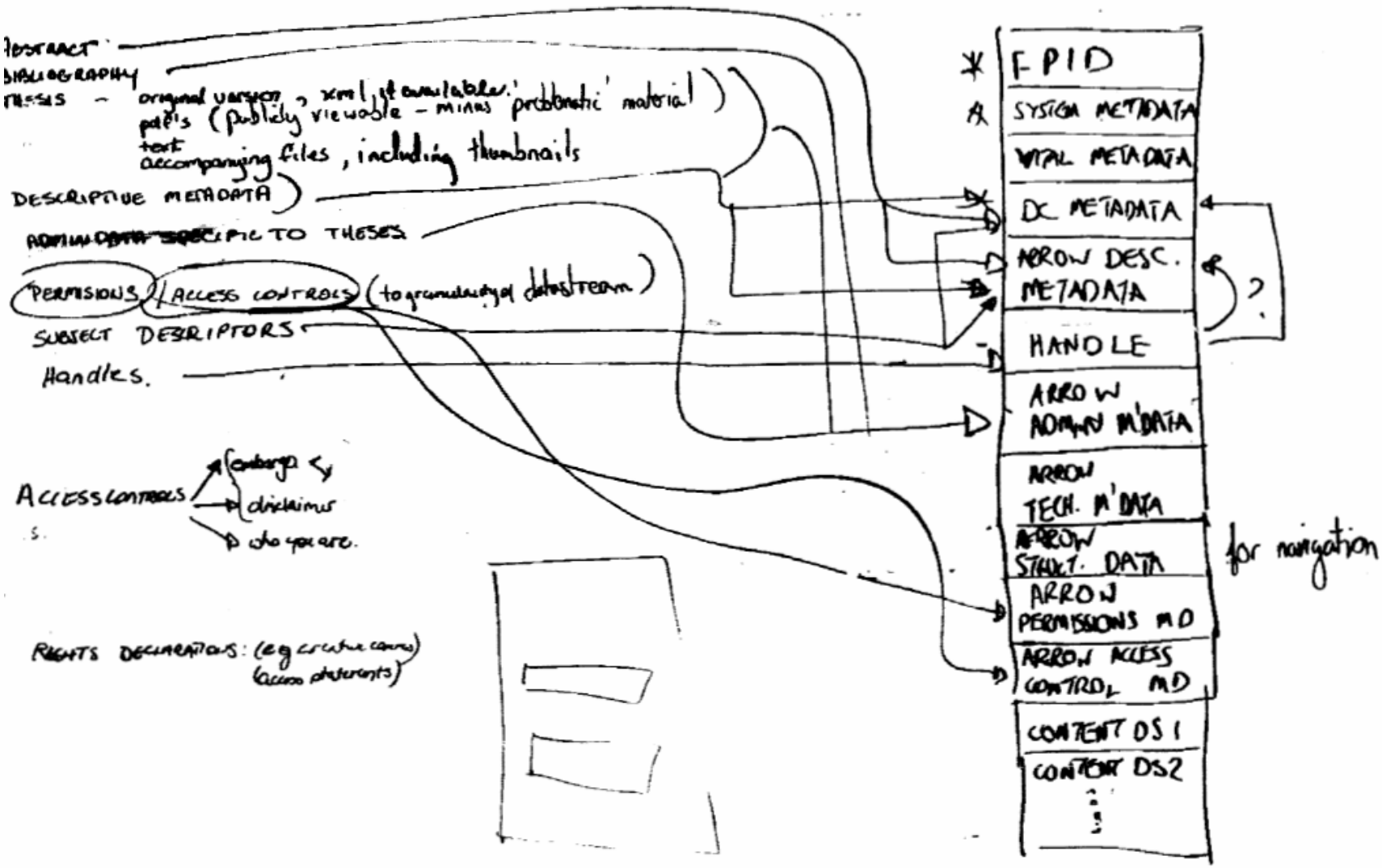
- Handles software has proved fairly painless
- Handle assignment in ARROW software has taken a while to get to the original vision
- Haven't yet needed to dis/re-aggregate content
- Persistent Identification and Linking Infrastructure (PILIN) project is currently looking at suitability of handles for a national Australian persistent ID system
- Some debate within ARROW on need for persistent identifiers at all

Content Model Decision

- Lots of early discussion about how to model different types of content objects
- Atomistic approach
  - each object has one datastream
  - use some mechanism to explicitly link together related objects
- Compound approach
  - objects can have multiple datastreams of differing types
  - linkage of datastreams happens implicitly
- Arrow chose to use the Compound object model

arrow
australian research repositories online to the world

From Whiteboard...



To Model...

Fedora PID
Handle
DC Metadata
VITAL System Metadata
DS 1 (Thesis Abstract)
DS 2 (Thesis Full Text)
DS 3 (Accompanying video)
DS 4 (Accompanying dataset)

Content Model Retrospective

- Has simplified software development
- Has complicated matters with respect to metadata
  - each object component can't easily have its own metadata
- Combination of RELS-EXT and RELS-INT may offer a more sustainable solution
- Still not sure that this was the right decision (or **if** there is a right decision)

Consortium Decision

- Consortium comprises Monash University (lead institution), National Library of Australia, the University of New South Wales, and Swinburne University of Technology
- Designed to incorporate small and large institutions, as well as the searching and indexing expertise of the NLA
- Each partner has two members on the ARROW Management Committee

Consortium Retrospective

- Mostly a positive experience
- Having partners brings in lots of ideas, and helps to share the load
- Different partners have expertise in different areas
- However:
  - Different priorities at the partners has created tensions
  - Keeping all the partners on the same page is not easy

Development Model Overview

- Entered into a partnership arrangement with VTLS Inc. (www.vtls.com)
- VTLS provided:
  - Development expertise and staff
  - Infrastructure for future support and development
- ARROW provided:
  - Intellectual property and design specifications - Versions 2 through 4 of VITAL largely based on ARROW input
  - “Real world” use cases, testing and feedback

Development Model Rationale

- Time to market – hopefully faster than doing it ourselves
- Focus on defining what we want, rather than trying to develop it ourselves
- Sustainability once project funding ends
- Offering a vendor for members to turn to for support and advice

Requirements of successful development

- Need for tight specs to:
  - Define work to be done
  - Avoid misunderstandings
  - Make agreement on successful completion of milestones easier
- Lots of communication
  - Distance and time zones a constant issue
  - Weekly teleconferences
  - Mailing lists
  - Wiki of documentation
  - Face to face meetings at least every 6 months

Development Retrospective

- Development feels slower on the inside
  - Delays caused by:
    - Some things being harder than they initially seemed
    - Staff turnover
    - Multiple players (ARROW, VTLS, Fedora, other organisations)
      - Fedora 2.0 => 2.1 delay
    - Scope changes requested by consortium partner
- Different priorities between ARROW and VTLS

Development in 2007

- Open Source building on Fedora
  - Sustainability of this development?
- Interoperability of Open Source with VITAL
  - How can this be managed effectively?
  - How do you keep versions in sync?
- Commissioning of specific requirements from VTLS

Development in 2007

- ARROW Mini/Partner Projects
  - Funding OS work on specific needs at ARROW partners and members.
- Developing incremental functionality
  - RM4 (integration with Research Management tool)
  - VALET (web-ingest tool)
  - OCLC (metadata interoperability)
  - BRACER (access control creation and management)

General Conclusions

- It's been a fun ride!
- Still very early days for institutional repositories as we all work out how to embed them into our institutional fabric
- ARROW now starting to focus on what to work on (and how) post the project funding

arrow
australian research repositories online to the world

Questions?

- Project Manager
  - David.Groenewegen@its.monash.edu.au
- Technical Architect
  - Andrew.Treloar@its.monash.edu.au