UNSW LIBRARY

Becoming the authoritative source taking repositories centre stage

Howard Amos

Taking repositories centre stage

- What do we mean by repository
- Why are we bothering
  - What services are required
- The environment they operate in
- A look at UNSWorks
- Where to next

What's a repository

- *Set of services for the management & dissemination of digital material*

  (Lynch 2003)

- OA movement – where does that fit
- What about "open science"/"open data"
- How wide is our community
- How does it support the academic mission
- How does it support the researcher
- Where does the library fit

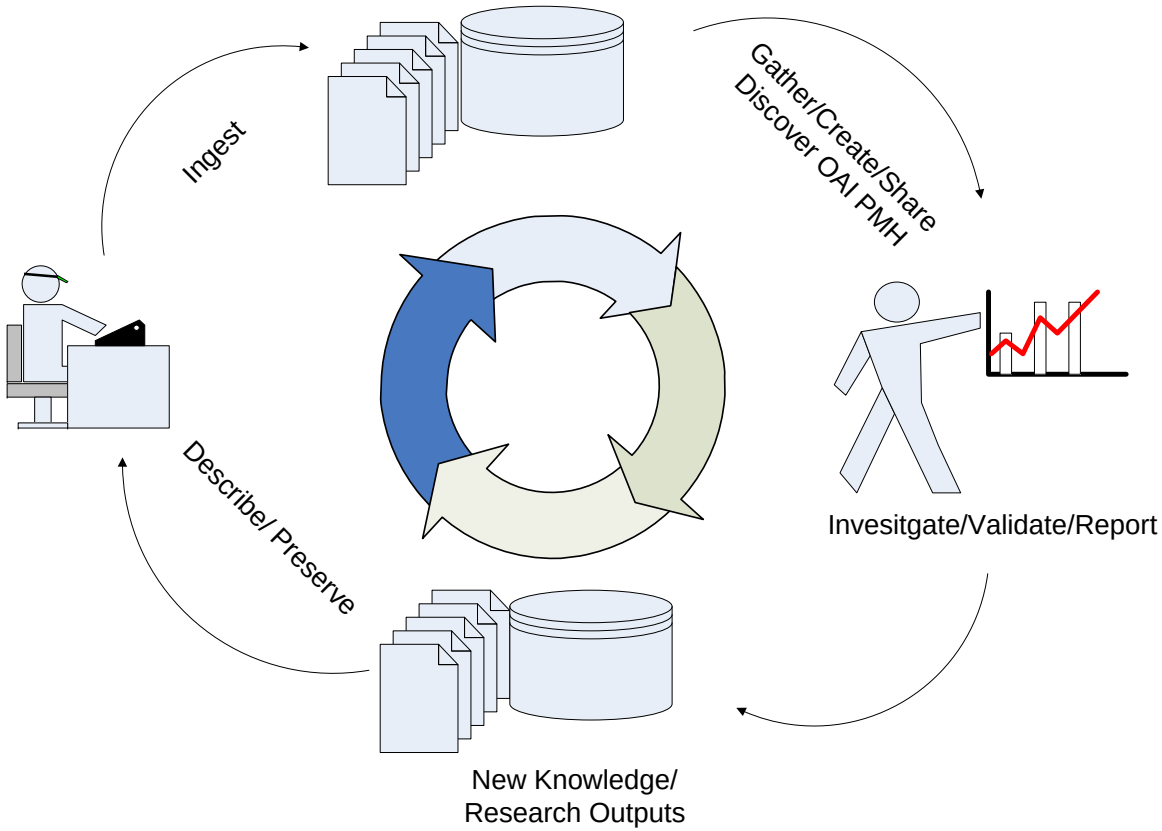
What does it need to do

- Hold research outputs
- Hold research data sets
- Make these discoverable
- Make them appropriately accessible
- Facilitate the use and re-use of material
- Integrate with the research process

What does this mean

- Not just familiar digital library services
  - It’s also about making research data accessible
- Facilitate the use & re-use of material
  - Value grows as non-anticipated use of datasets grows
- Skills needed
  - Data management
  - Access controls
  - Describe for discovery

How it might work



UNSWLIBRARY

How is it going to do this

- Core functions of an institutional repository
  - Material submission
  - Metadata application
  - Access control
  - Discovery support & Distribution
  - Preservation
  - Reporting

Core functions

- Material submission
  - Needs to be easy
  - Self submission as much as possible
  - Auto populate and control terms
  - Allow for common workflow
    - Submission on behalf of researchers

Core functions

- Metadata application
  - Transparent/ low intervention
  - Assist with identifying and applying schemas
  - Adhere to standards with validation
    - (MARCXML MODS DC DDI MatML etc. etc.)
  - Ease of maintenance

Core functions

- Access control
  - Meets local and federated needs
  - General and specific rules
  - Standards based SAML XACML etc

Core functions

- Discovery
  - Harvestable (OAI)
  - Google able
  - ORE able
  - Push
    - to subject based repositories
    - Research centre/researchers web sites
  - Pull
    - Other repositories/resources (Citeseer PubMed etc.)

Core functions

- Preservation
  - Describe object formats consistently
    - JHOVE
  - Ensure can still access them
    - AONS
  - Build and support persistency
    - Provide an enduring home

Core functions

- Reporting
  - Internal admin
    - Research census/strengths/competitive advantage
  - External reporting
    - HERDC (Aust.)
  - For the researcher
    - Part of the promotions process
    - Consistency of output/expose research depth
    - Contribute to bibliometric measures
    - Attract other researchers & HDR students

Who’s responsible anyway

- Librarians aren’t the drivers
  - We can contribute to defining needs
  - We support scholarly communication and the academic mission by supporting the research process
- Assist in devising/extend schemas to maximise discoverability
  - Enabling open science/open data
- Provide expertise in collection management
  - Data definitions for resources
  - Management of descriptive consistency
  - Data clean up/normalisation

Context is king

- Repository should be community-driven and community-focused
- To be sustainable needs community engagement
- To thrive needs to replace/improve existing admin services not add to them
- Could be institutional, knowledge area or cohort based

But it must be….

- What our Researchers find useful
  - Stewardship of material (including data sets)
  - Provide efficiencies
    - Gather once use many
  - Showcasing community member's work
    - Extend and expand exposure
    - Facilitate digital scholarly communication
  - Push material to where they want
  - Measure and report on content or usefulness
  - Integrate with other research tools & processes

There is a cost….

- What will our paymasters find useful
  - Ease of administration
  - Improve the University's profile
  - Remove duplication of effort
    - One place for research publications
  - Capture savings
    - Efficiencies from centralisation
  - Improve reporting and accountability

What else

- Making it accessible means managing it
  - Research data and data sets
  - Data a primary source material
  - Data as part of compound publications
- Managing these requires
  - New ontologies & consistent description
  - Cross discipline identifiers
  - Data structuring conventions
  - Search & retrieve protocols
- Cross border administration - a whole new ball game

Moving from the general to the specific

UNSWorks

An initiative of the UNSW Library

UNSWLIBRARY

Australian environment

- Increased Accountability
  - HERDC
  - RQF/ERA
- Developing infrastructure for e-research
  - NCRIS
  - Platforms for collaboration
  - ANDS

UNSW Environment

- Strategic directions
  - Support collaborative research
  - Develop e-research capabilities
  - Improve bibliometric measures
  - Integrate administrative infrastructure
    - Remove duplication
  - Library's role:
    - provide UNSW researchers with a competitive advantage using digital tools and services

Beginnings

- 2004 the ARROW Project
  - With Monash, Swinburne and the National Library
- Now with 15 more Australian universities
- UNSW contributed to:
  - Standards
  - Workflows
  - Software solutions

UNSWorks development

- Development strategies
  - Clear well defined scope and purpose
    - For RESEARCH outputs
  - Maximise interoperability with other UNSW systems
    - Research Office
    - IARO
  - Ease of use/efficiencies
    - Capture one use many
  - Standards based architecture

UNSWorks requirements

- Priorities
  - HERDC/RQF/ERA
  - ADT upgrade
  - Working papers/research series
    - Supporting digitisation projects
  - Reduce reporting overheads for researchers
  - Build useful functionality
    - Bulk ingest
    - Importing data sets

UNSWorks

- Implementation
  - Expertise:
    - Employed a BA (important) and a developer
  - Played with lots of different types of content
    - Some we now are trying to lose
  - Identified different individual needs
  - Part of wider ARROW community
    - Network of implementers

UNSWorks

- What next
  - Support for major UNESCO Project
  - Dataset management
    - National Centre in HIV Social Research
  - Complex items
    - Research outputs combined with data
  - Access
    - AAF
    - ORE

The way forward

- Build skill sets
- Help define architecture/service framework
- Help identify standards and schema
- Work closely with researchers
  - Participate in the research process
- Break down Silo activities
- Define border responsibilities
  - Deploy/Build tools

Some other questions

- What infrastructure do we need
  - Who will design/build/run this
- What collection management practices do we need
  - What to keep, in what form, using what tools?
  - Selection, weeding, destruction etc.
- How do we preserve this new scholarly record