Managing local institutional assets

OAI on campus

Chris Rusbridge, University of Glasgow
Contents

• OAI: disclosing campus assets
• ePrints and eTheses
• JISC FAIR Programme
• Daedalus
Open Archives Initiative 1

• Grew out of specific proposals for ePrints (Santa Fe agreement?)...
• Generalised to mechanisms for disclosing metadata from data providers to services
• Widely applicable...
  – Images
  – Video clips
  – Audio resources
  – Learning objects
  – Union catalogues
  – Perseus Digital Library...
Open Archives Initiative 2

• First version of protocol released last year
  – Many implementations and uses
  – Particularly being applied to disclosing existence of assets you want publicly known

• Second version of protocol this year
  – Few implementations yet

• General model is slurping metadata into union DB...
OAI model

Data provider 1

Data provider 2

Service provider
OAI model

Data provider 1

Data provider 2

Service provider
OAI model

Data provider 1

Service provider

Data provider 2
OAI model

Data provider 1

Data provider 2

Service provider
Z39.50 model

Target 1

Service provider

Target 2
Z39.50 model

Target 1

Service provider

Target 2
Z39.50 model

Target 1

Target 2

Service provider
Differences…

- **Z39.50**
  - in real time (failures, limited processing, etc)
  - Multiple metadata mappings acting as semantic filters

- **OAI**
  - Results based on harvesting in past
  - Metadata presented as provided
ePrints

- Precursors
  - Los Alamos Pre-print archive
  - NCSTRL
  - WoPEc
- Now generalised using OAI...
- Subversive proposal...
ePrints questions

• Copyright
  – Institution versus author ownership
  – assignment of rights to publisher versus limited licence
    • Few now refuse ePrints permissions
  – ‘Prior’ publication if pre-prints mounted?
    • Ingelfinger rule; few now invoke
ePrints questions

• Who should be the data provider?
  – Institutions?
    • Self sustaining, self interest
    • Multiple authors across institutions??
  – Specialist sites (eg ArXiv)
    • Business case problem
  – Subject organisations?
    • Business case problem
ePrints questions

• Who mounts the paper?
  – Librarians? Doesn’t scale
  – Authors? Too busy
  – Clerical staff? Requires commitment, management, etc

• Who provides the metadata?
  – See above!
ePrints issues

- Technical (getting solved)
- Metadata quality & consistency
  - Author-provided keywords
  - Dataset structure: divisions on subject lines?
  - Index content via web? (DP9)
- Collection policy (published vs working papers)
- Usefulness of services...
  - Critical mass of papers missing
  - Advocacy needed
- Long term preservation
ePrints dimensions

- Our life sciences faculty has 4,753 published papers in its RAE publications DB, since 1995 (RAE submission ~480)
  - GU RAE submission about 4,400 papers...
  - We have 33 papers in our ePrints server!

- Alan Kent’s harvester stats show:
  - Glasgow 17
  - Nottingham 41
  - RDN 388
  - UKOLN 113
  - Grand total 780,000
ePrints implementation

• Freely available data provider software
  – eprints.org and others

• Various service providers harvesting metadata, including
  – ARC
  – OAIster
  – my-OAI
  – Perseus
  – RDN
Faculties

- Arts (3)
  - HATII (1)
  - Music (1)
  - STELLA (1)
- Biomedical & Life Sciences (3)
  - Division of Environmental and Evolutionary Biology (3)
- Computing Science, Mathematics & Statistics (1)
  - Computing Science (1)
- Divinity
- Education
- Engineering
- Law & Financial Studies
- Medicine (including Dentistry and Nursing)
- Physical Sciences (1)
  - Physics (1)
- Social Sciences (1)
  - Psychology (1)
- Veterinary Medicine

Administrative Units

- Central Administration
- Information Services (21)
  - Archives (1)
  - Computing Service (4)
  - Library (12)
    - GAELS Project (3)
eTheses

• Networked Digital Library for Theses and Dissertations
  – Virginia Tech, Ed Fox
  – 145 members worldwide
  • City University
  • British Library
  • Edinburgh
  • Glasgow
eTheses

• Universities Thesis Online Group (UTOG)
  – Originally focused on digitisation of print theses
  – Commissioned studies via Edinburgh

• Dimensions
  – 370 PhD theses defended each year at Glasgow
  – VT has >3K ETDs, >3M PDFs supplied in one year!

• Issues
  – Regulations
  – Authenticity
  – Restrictions
  – Preservation
JISC FAIR Programme

- JISC Circular 1/2002: Focus on Access to Institutional Resources
  - ‘inspired by the vision of the Open Archives Initiative [that] digital resources can be shared between organisations’
JISC FAIR Programme

- Large no of bids
- 14 projects funded
- 5 clusters
  - ePrints (4)
  - eTheses (3)
  - IPR (1+)
  - Institutional portals (2)
  - Museums and images (4)
Daedalus

• ‘development of a network of different data providers for a range of institutional assets to be exploited by a local harvesting service’
  – Academic [Published] papers
  – Pre-prints and Grey Literature
  – eTheses
  – Research Resource Finding Aids (Archives and Special Collections)
  – Institutional Management, Policy and Regulatory documents (linked to CDocS)
Daedalus

• Local service provider
• Advocacy
  – Eg http://www.gla.ac.uk/createchange
• Programme of studies, workshops and reports
• Links to SHERPA, HAIRST, eThesis projects…
Summary

• OAI approach to sharing metadata has great potential
• Barriers more human than technical
• Needs concerted advocacy
• Impact of lower quality metadata unproven