Oxford Digital Asset Management System (DAMS) Update

Neil Jefferies
R&D Project Manager
Systems & eResearch Services (SERS)
Oxford University Library Services (OULS)
Agenda

• Overview
• Fedora-Commons
• Honeycomb/ST5800
• Mass Digitisation Ingest Components (MDICS)
• Projects
Overview

- Oxford Research Archive (ORA)
  - ePrints, eTheses, Working/Conference papers
  - Harvesting from PubMed, ARXIV, RepEc...
  - http://ora.ouls.ox.ac.uk

- Other Applications due to come online
  - Digitised Books
  - Electronic Ephemera
  - Manuscripts
FEDORA-Commons

- Loosely coupled, componentised system
  - See Ben's presentation for later
- Heavy use of other existing OSS projects
  - Unmodified
- VMWare ESX provide a flexible environment
  - Server put-up/tear down to suit load
    - e.g. Fewer indexers & more virus scanners/text extractors during mass ingest
  - Provides resilience to all components
Honeycomb/ST5800

- One is production use in DAMS
- One is a pre-ingest store
  - Electronic Ephemera project
  - DIAMM (dark)
  - Will migrate to Thumper in due course
- We need more...
  - But have physical space/power issues
- Working with Sun to develop metadata functions
Mass Digitisation Ingest Components (MDICS)

- Ingest of Google Library Project output into DAMS
  - Download packages
  - Unpack and validate contents
  - Generate metadata (descriptive from LMS)
  - Generate delivery formats (PDF?)
  - Multiple parallel streams on T2000's
- First version being performance tested
Projects

- JISC has been a key driver in the UK
- Interoperability between systems
  - Systems integration beyond repository
- Preservation
  - Maintain diversity of tools
- Expand the scope of the repository
  - New content types
eTheses

- Oxford has an electronic thesis mandate
- Based on interoperability and sharing
- DART Europe (http://www.dart-europe.eu)
  - European portal based on OAI-PMH harvest
- Nereus/NEEO (http://www.nereus4economics.info/)
  - European Economics focused
- ETHOS (BL) (http://www.ethos.ac.uk/)
  - UK Wide with Digitisation
Research Excellence Framework

- Government HE performance review
  - Determines baseline funding for institutions
- Heavily focused on bibliometrics
- Oxford Research Archive is a logical place to do some of this
  - Integration with bibliometric sources
  - Integration with Research Services systems
Bridging the Interoperability Divide
(http://www.sers.ox.ac.uk/bid)

• Links three Oxford repository systems
  – OULS: Oxford Research Archive (Fedora)
    • Creates objects linking to foreign content
    • RDF captures/exposes relationships
  – OUCS: Virtual Learning Environment (Sakai)
    • Metadata harvest + Archiving
  – OeRC: Grid Datasets (SRB)
    • Metadata harvest
PRESERV2
(http://preserv.eprints.org)

- Southampton, The National Archive
  - Consultants: British Library & Cambridge
- Preservation tools/services for repositories
  - API's for integrating tools with repositories
  - Feeds into PLANETS via TNA
  - Covers EPrints, FEDORA and D-Space
- OR08 – CRIG Challenge
  - Migrate between ePrints <-> Fedora via OAI-ORE
SHERPA DP2
(http://www.sherpadp.org.uk/sherpadp2.html)

• Many institutions plus Arts & Humanities Data Service
• Examining model where AHDS provides a centralised preservation service
  – Accepts feeds from EPrints, FEDORA & D-Space
• We expect to be a central provider in the long term
Forced Migration Online
(www.forcedmigration.org)

• Repository service provision model
  – We provide storage and object management
  – They develop a custom application layer
  – We influence design to enhance re-use

• Content is thematic and contains non-Oxford material
  – Suitable Oxford material automatically included in ORA
DISC-Datashare
(http://www.disc-uk.org/datashare.html)

- Edinburgh, Southampton, Oxford ...
- Bringing datasets into repositories
  - Focus on smaller datasets (Social Sciences)
  - More varied standards and formats
    - How to ensure re-usability
- Metadata standards are a major issue
  - Often domain specific – DDI
    - And lots of it!
  - Diversity of structures hard to capture
PARADIGM/CAIRO
(http://cairo.paradigm.ac.uk/)

- Ingest and preservation of complex objects
- Personal digital effects – PDA's, mail files, etc.
  - Common workflow with physical effects
  - Wide variety of formats / technologies
    - “Digital Archaeology”
  - Snapshots during lifetime
  - High security: loss preferable to premature release
- Captures complex relationships
Digitisation Projects

- Oxford Digital Library (http://www.odl.ox.ac.uk)
- Google Libraries Project
- John Johnson Ephemera
  - Variety of formats, aggregations
- Integration with Imaging Service
  - Paid digitisation of any material
- European Digital Library (http://www.edlproject.eu)
  - Portal to National cultural heritage collections
Questions

Neil Jefferies
neil.jefferies@sers.ox.ac.uk
www.sers.ox.ac.uk

Oxford Research Archive
ora.ouls.ox.ac.uk

Developer's Blog
oxfordrepo.blogspot.com

Google Code
look for: python fedora-commons