Oxford DAMS – Design for Longevity

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

- Introduction to SERS/OULS
- DAMS Design Principles
- DAMS Architecture
- Projects
Introduction to SERS/OUULS

- Bodleian one of ~110 libraries
- 32 OUULS Libraries
- 29 Non-OUULS Libraries
- 46 College Libraries
- And a few others!

OUULS

- 750 Staff
- £25M budget (€32)
- 11 million items
- 156 shelf miles (250 km)
- SERS provides all electronic services
Core Services

- Library Management System
- Electronic Resources
  - 600 Bibliographic Databases, 40K Journals
- Digital Library
- Hybrid Library
  - Integrated discovery and delivery
  - Integration beyond library/institution
  - 365x24x7
Digital Library

- Digital Asset Management System (DAMS)
  - Common infrastructure for DL applications
  - Legacy digitisation projects (~5TB)
  - 1M digital surrogates (~5TB)
  - “1M” Google Library Project (?TB)
  - Current digitisation programmes
  - Born-digital materials
DAMS Design Principles

• The Theory...
  – Content centred design
  – Layered component architecture
  – Generics over specifics
  – Capability focus
• Hindsight is a wonderful thing...
Content Centred Design

• “Librarian's View”
  – People change
    • Requirements & processes change
  – Buildings/containers change
    • Only the content persists
• User requirements/process modelling can only take us so far
  – End up as limitations
Layered Component Architecture (1)

- Systems and standards change rapidly
- "Big bang" implementations are a bad fit
  - Infrequent, inflexible, failure-prone
- Loosely coupled components
  - Web services model a good fit
- Any component can be replaced
  - Design/select with a view to replacement
- Few, simple, well-defined API's
Layered Component Architecture (2)

- Multiple layers
  - Applications (Collections, Archives)
  - Tools (Indexing, Discovery, Scheduling)
  - Object Management
  - Storage
- Architecture requires discipline
  - Concise, simple, standardised API's
Layered Component Architecture (2)

Storage
- Honeycomb
- Thumper ZFS
- HFS

Object Management
- Fedora
- Fedora
- Fedora

Tools & Services
- Search
- Reporting
- Scheduling
- Job Mgt
- Virus Scan
- Text Extract

Applications
- ORA
- Digitised Books
- Ingest

April 2008
Digitalna Kniznica 2008
Generics over Specifics

• Identify generic operations in applications
  – Push into the tools layer
  – Simplifies future application development
  – Replacement can upgrade all applications

• A generic capability is much less likely to become obsolete than a specific one
  – Protects investment in development
Capability Focus

• Every design decision should be reviewed...
  – What does it allow us to do?
  – What might it prevent us from doing that we will really regret later?
• Scalability will always be an issue
• Short workflows
  – Decouple actions as much as possible
DAMS Architecture

• The practice...

Oxford University Research Archive
ora.ouls.ox.ac.uk
'Backend' systems - service provision view

Service Providers (outwardly focused)
- Darwin Calendar Server
- RDFLib Triplestore
- Solr (Search)
- eXist (XML db)

Preservation and maintenance services (inwardly focused)
- Event Queue
- Text Extraction Service
- Malware Scanner(s)
- Metadata synchronisers
- Email/XMPP alerters
- Message Logger

FEDORA

Sun Honeycomb
Frontend to Backend connections

'Backend' systems
- Solr
  - HTTP Query
- Calendar Server
  - CalDAV
- RDFLib (RDF repository)
  - SPARQL Query
- eXist
  - HTTP XQuery
- ActiveMOS (IMS 1.1)
  - OpenWire
  - Scalos
  - Jave IMS
- FEDORA
  - REST
  - APLM
  - APLA
  - SOAP
  - APLM
  - APLA
  - Tripoli
  - RINergy
  - PROAI
  - iGAL

Frontend Services - "ORA"
- HTTP Client
- CalDAV Client
- SPARQL Client
- Reports
- RSS Feeds
- OpenWire Client
- SOAP Client
  - Bookmark Manager
- HTTP Client
- SOAP Client
  - Object Display
  - Object
  - Authorization Module
  - GAI-PMI
  - Authentication Module

Frontend Ingest & Edit
- HTTP Client
- SOAP Client
- HTTP Client
- SOAP Client
- Combine Harvester
  - Automated Import/Export
  - Object creation and editing
- SoPAC Client
- CalDAV Client
Example Design Decisions

- Fedora Commons (www.fedora-commons.org)
  - Flexible object/metadata model
- UUID's preferred over handles
  - Scalability, dependencies
- No OAIS/METS!
  - Longevity, efficiency, scalability
- Calendaring
  - Generic capability, additional benefits
Projects Themes

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

- BID (Bridging the Interoperability Divide)
  - Oxford Research Archive (Fedora)
  - Virtual Learning Environment (Sakai)
  - Grid Datasets (SRB)
- DART Europe, NEEEO, ETHOS
  - eTheses projects based on interoperability and information sharing
  - Different scopes and requirements
Preservation

• PRESERV2
  – Southampton, The National Archive
  – Preservation tools/services for repositories
  – OR08 – Whole repository migration

• SHERPA DP2
  – Multiple UK Institutions and Arts and Humanities Data Service
  – Provision of outsourced preservation service
Expansion

• DISC Datashare
  – Bringing datasets into repositories
  – Metadata standards
• PARADIGM/CAIRO
  – Complex objects, personal digital archives
  – Security and longevity are key
• Digitisation: Google, John Johnson...
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

April 2008