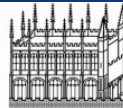


Hydra at Oxford

Neil Jefferies

R&D Project Manager, Bodleian Libraries

Neil.Jefferies@bodleian.ox.ac.uk



Bodleian Libraries
UNIVERSITY OF OXFORD



UNIVERSITY OF
OXFORD

An important bit of background...

We don't really use FEDORA

- FEDORA **was** a strategic choice in 2007
 - VTLS Vital to provide institutional repository (ora.ox.ac.uk)
 - Project failed, ORA public interface subsequently rewritten from scratch
 - VTLS Valet (open source) continues to provide deposit workflow
 - ORA the only digital content system still based on FEDORA
- What we liked about FEDORA
 - Flexible generic object model – multiple datastreams with versioning
 - Semantic model – RELS-INT/RELS-EXT
 - REST API
 - Storage abstraction
- What we didn't like about FEDORA
 - Unnecessary wrapping – FOXML not good for preservation of active objects
 - Active-use is the major economic justification for preservation
 - No, we don't like METS either!
 - Lack of modularity – external triple-store grief
 - Feature bloat – content models, XACML etc. - made worse by lack of modularity
 - Installation (need I say more)
 - Silent periods

CDL Microservices

- CDL “What's the minimum amount of code you need to add to a filesystem to make it look like a repository?”
 - Pairtree, Namaste, BagIt
 - Pluggable Web Services provide bulk of repository/preservation functionality
- Ben O'Steen “What's the minimum amount of code we need to add to a Microservices repository to make it look like a FEDORA repository”
 - Enough like a FEDORA repository to do the things we like
 - But retaining the scalability and modularity of the Microservices model

...rather less than we thought!

- Databank came into being in 2008
- **2010-12** JISC/HEFCE (UMF) funded [Admiral/DataFlow](#)
 - Prototype/productionise DataStage/DataBank
 - Libraries, Computing Services, OeRC, IBRG, UKOLN, Canonical
 - Lightweight data management/archiving

DataBank

DATABANK search data packages Search

Silo ● Data packages ● File names ● Any ●

home browse about Login

silo: ww1archives > data package: gwa-2156 >

Data package gwa-2156

The Autograph Book of Beryl Ellis (1)

Version 2

You are currently viewing the latest version (2) of the data package

View other versions:
0 1 2

Access Information

Data package is openly accessible

Data package contents

2162.jpg (115.4 kb)

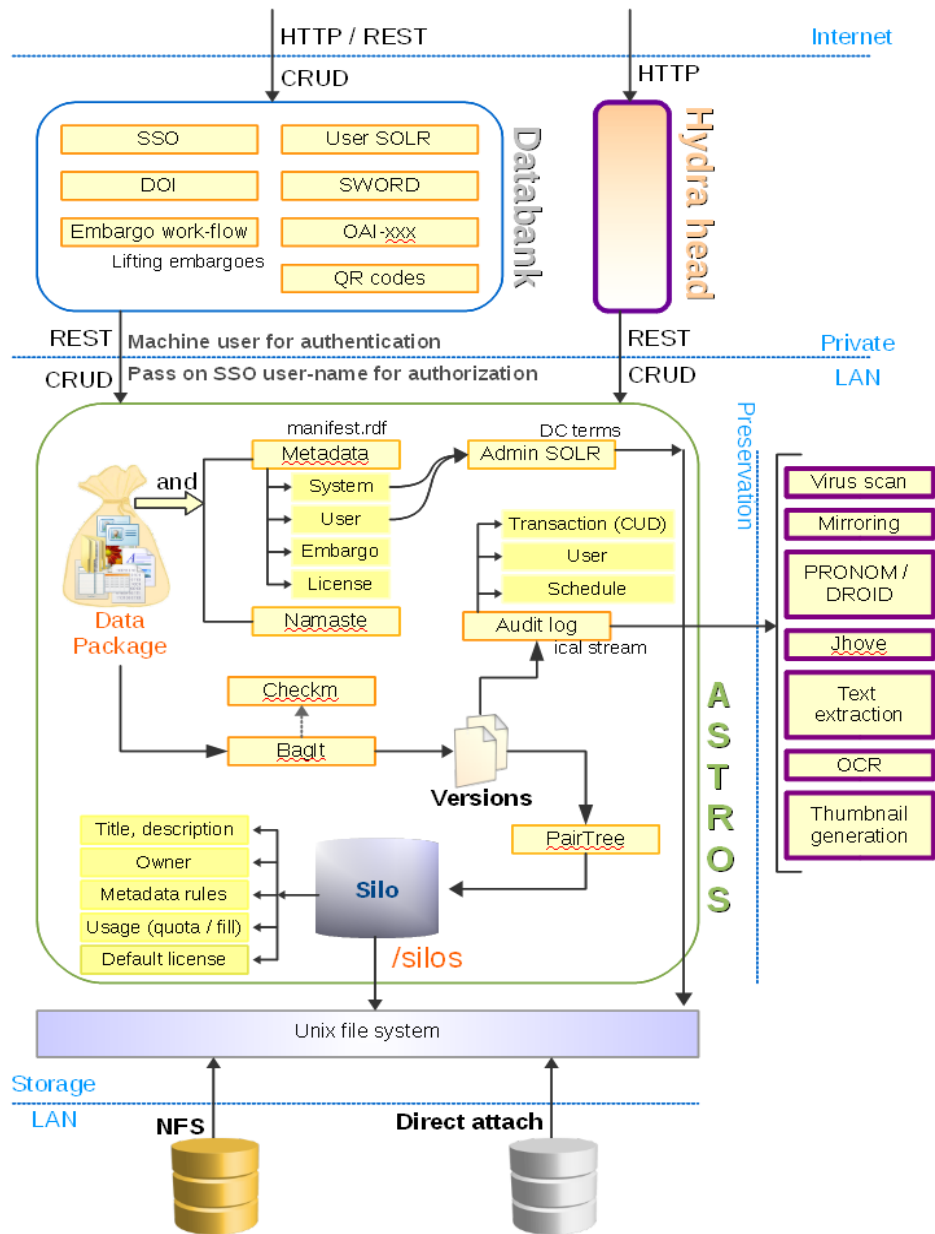
manifest.rdf (2.8 kb)

Metadata

Title	The Autograph Book of Beryl Ellis (1)
Identifier(s)	GWA_0066_001.jpg gwa-2156
Description	Beryl Ellis was a nurse at Moor Green Lane Hospital, Birmingham, during the Great War. This autograph book contains pictures and poems drawn and written by Veterans of the First World War she cared for at the hospital.
Created	December 04 2011, 08:59 PM January 01 1917, 07:18 PM
License	CC0 1.0 Universal (CC0 1.0). See http://creativecommons.org/publicdomain/zero/1.0/legalcode
Mediator	ww1archives
Modified	December 04 2011, 08:59 PM
Publisher	Bodleian Libraries, University of Oxford
Rights	http://ora.ouls.ox.ac.uk/objects/uuid%3A1d00eebb-8fed-46ad-8e38-45dbdb4b224c
Is embargoed	No
Current version	2

- Bodleian Data Repository (in dev since 2008) parallels ORA
- “Data” currently defined as “Research outputs that don't fit in ORA”
- File and metadata format agnostic
 - supports packages (zip & tar)
 - component subaddressing
- Built on “FEDORA-Lite” object model
- Assigns DataCite DOI's
- Manages embargos
 - Secure, dark archive is segregated
- Manual and SWORD2 deposit
- REST API
- Debian Packages or OVF

Databank and Astros Architecture



Architecture

- Microservices orchestrated using message queues
 - Event streams act as schedule, log and provenance
 - Queues can be exposed externally
- Search/browse interface
 - SOLR built-in
 - REST API provides meaningful responses to requests for text/html
- Multi-streamed RDF Metadata
 - Segregated by type/accessibility
 - Other XML metadata also supported
- Split authentication/authorization
 - Systems integration
- Unix file system semantics
 - Less abstraction?
 - LTFS

So...why Hydra?

The subset of FEDORA functions implemented by DataBank/ASTROS almost exactly matches the subset used by (most) Hydra heads. This subset, in essence, characterises a generic semantic object store. We can use this for pretty much everything...

Hydra in the near future

- Replace VTLs Valet for ORA ingest
 - Still running over FEDORA at this point
 - Migration off FEDORA still planned
- Consolidate legacy digitized materials
 - Migrate into ASTROS from many scattered servers/websites
 - Publish through Digital.Bodleian (Armadillo)
 - Metadata is somewhat variable MARC->MODS->METS
 - Need a MODS editor
- Archival materials
 - Physical, hybrid and electronic
 - EAD is problematic
- Shared Canvas/IIIF
 - Viewer encapsulated as a Hydra head
 - Annotation/transcription tools

Catalogue 2.0

- Objects have a context from which much of their meaning is derived
 - Include context objects representing people, places, events etc.
 - Catalogues become contextual skeletons fleshed-out by “traditional” digital objects
 - Authority lists become prosopographical and biographical resources
 - Geopolitical and temporal information
 - Aggregations become a key structural element
 - Should reflect actual knowledge – conflict with cataloguing practice!
- Annotations and files can be attached to any object
 - Context objects hold content...what is the difference?
- Objects are not static – preservation challenge...and benefit!
- This model works for almost any content from any period

Agents (Who)

Organisation

OpenOrg

Person

ARPRO

Location (Where)

FOAF

Physical

UNFAO

Geopolitics

Object (What)

MODS/DCterms

Work

Manifestation

Collection

Event Streams (When)

Organisational History

Membership/Projects/Activities

Biography/Geneology

Prosopography

Territorial Sucession

Historical Events

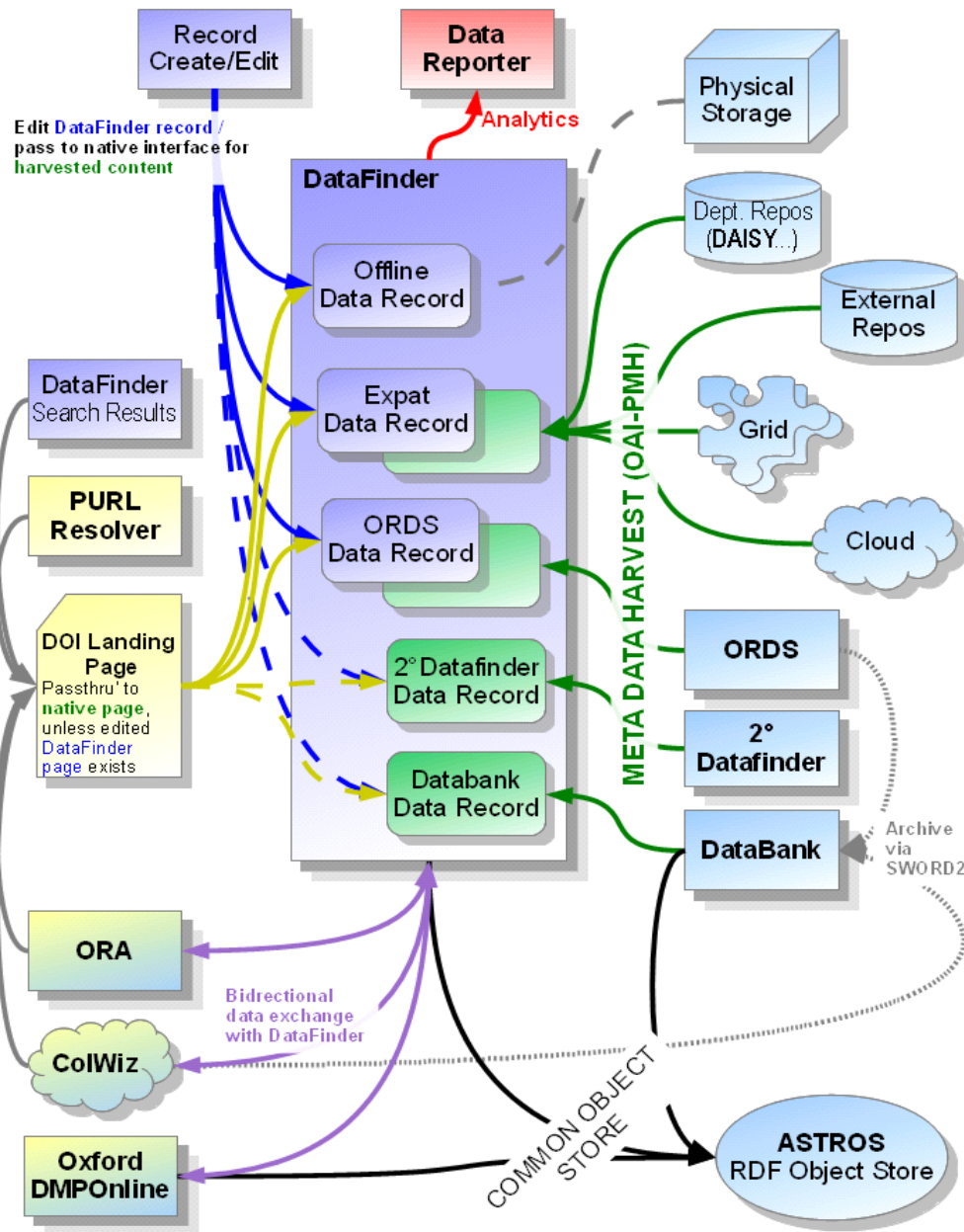
History/Provenence of Physical Object

History/Provenence of Digital Object

PREMIS Preservation/System Log

Scheduled Events (reviews etc)

Annotations



DataFinder

- Catalogue/registry of research data
 - Wherever and whatever it is!
 - OAI-PMH harvesting of external data stores
 - Manual record entry for non-electronic or non-harvestable data
- Search/browse interface
- DataReporter module
 - CERIF compatible
 - Analytics as well as content statistics
- Core Metadata schema based on DataCite
- Interfaces with many systems
 - "Hub" Of RDM activity
- Hierarchical architecture
 - Local catalogues, subjects specific or inter-institutional catalogues possible

Questions?

Neil Jefferies

R&D Project Manager, Bodleian Libraries

Neil.Jefferies@bodleian.ox.ac.uk



Bodleian Libraries
UNIVERSITY OF OXFORD