DMID Presentation

UK Fedora User Group

December 13 2010

Simon McLeish

Library of the London
School of Economics

Background

- LSE archives cover modern British political, economic and social history, the history of the social sciences, and the history of the LSE

- Digital collections are significant and growing, as are the requirements of their users

- Digital preservation is an increasing concern alongside our commitment to continue to develop innovative digital services for researchers and students

Our Digital Collections

Material Type	Existing Systems	Functional Requirements
Digitisation projects Images, Multi-page text, Maps (including geo-data)	Presentation through institutional CMS, custom Web sites and 3rd-party provision.	Rich multi-media Web and mobile interfaces Discovery, search and browse Preservation storage
Born-digital archives Any file formats and structures	Storage on original physical media and server (RAID, tape backup), Description in archives management system.	Preservation storage Integration with existing collection management Future integration with library discovery systems and multimedia presentation interfaces
Research outputs Mostly PDF, office formats Increasingly video, audio	Comprehensive collection management in mature EPrints repository.	'Traditional' IR functionality including cataloguing, ingest, discovery, presentation and harvesting to third parties

DMID

- The Digital Library Management and Infrastructure Development Programme (DMID) is a large scale cross-library initiative to build our capacity to collect, manage, preserve and provide access to our digital collections

- Aims to support all our digital archive content: born digital archives, digitisation projects, and research outputs

- Involves library staff from archives, information services and library IT departments

- Run by Ed Fay, who was to give this talk but was unable to attend this meeting

Foci

- The main concern is to provide a system which can:

  - accept digital material from depositors
  - securely hold that material without loss

- Then we can:

  - provide access mechanisms for the material
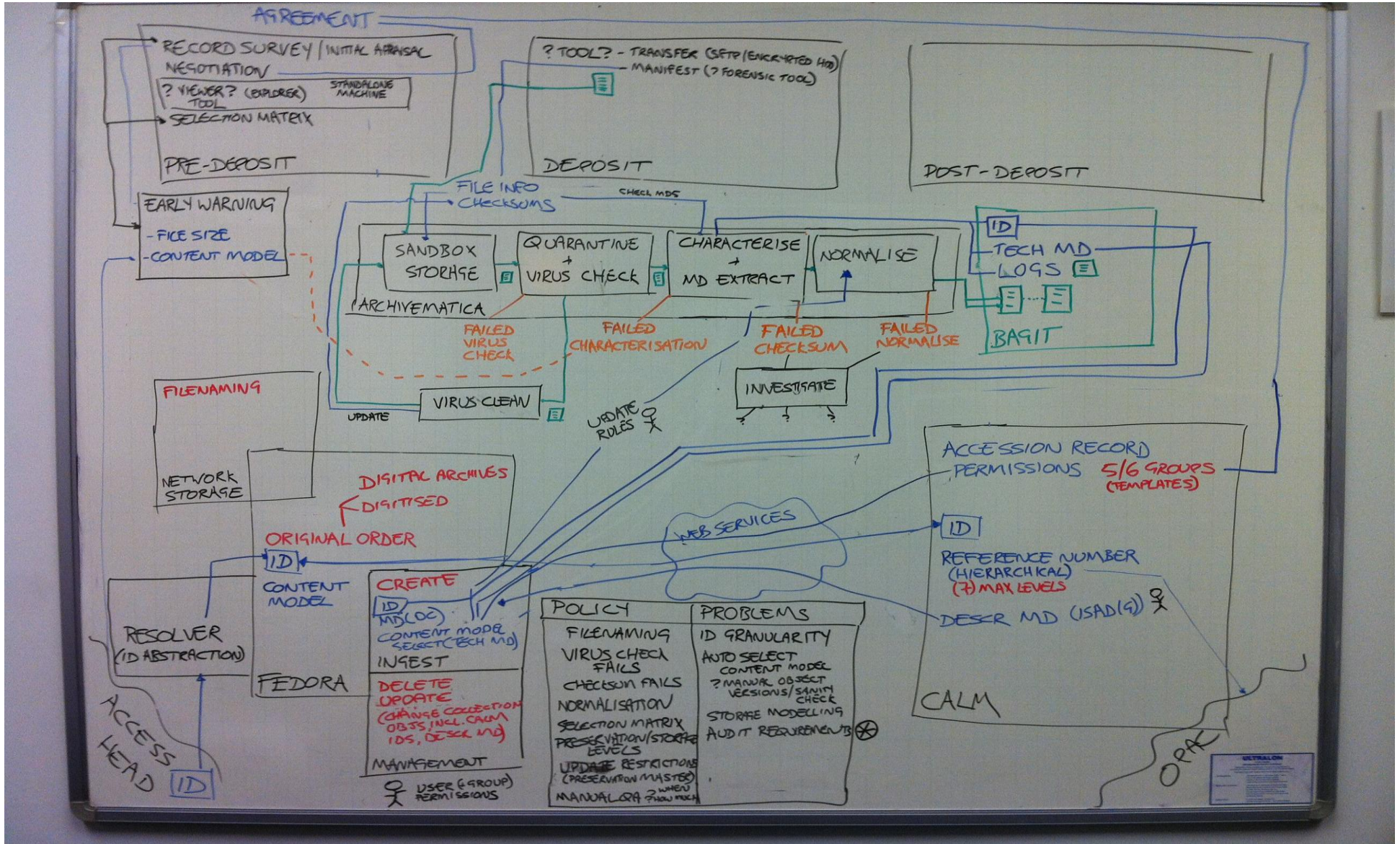 - preserve the materials for access in the future

Work So Far

- The initial phase of DMID has investigated our collections, their users and best practice in the wider community and produced functional specifications for testing against the current best-of-breed open source repository software

- Site visits to Oxford, Cambridge and Wellcome Institute

- Fedora felt to be best, essentially because easiest to mould to our requirements

- Ed contributed an article to Ariadne on the process we used to choose Fedora as our repository software of choice: http://www.ariadne.ac.uk/issue64/fay/

- Initial investigation into how best to meet our requirements in Fedora suggested parallels with Hydra and with the AIMS project, both of which involve Hull

- Two trips to East Yorkshire followed in autumn 2010

Use of Hydra

- Initial project timetable envisioned the development of required functionality from scratch

- Hydra project seemed to offer significant time savings – re-using other people's work!

- https://wiki.duraspace.org/display/hydra/The+Hydra+Project

- But there are some things it doesn't fully address...

Ingest Process

Collaboration with AIMS

- The LSE is not the only archive with issues around ingest

- The AIMS project looking at this as a collaboration with several universities

- Timetables not perfect match, but we should be able to work together

- Will share common tools such as Archivematica (or Rubymatica)

Preservation

- Currently thinking about preservation
  - How best to organise within Fedora
  - How this relates to Hydra etc.
  - How to create and update preservation metadata
- Advice would be useful...

Questions?