An Institutional Approach to Developing Research Data Management Infrastructure

Wednesday 8 December 2010

James A J Wilson, Michael A Fraser, Luis Martinez-Uribe, Paul Jeffreys

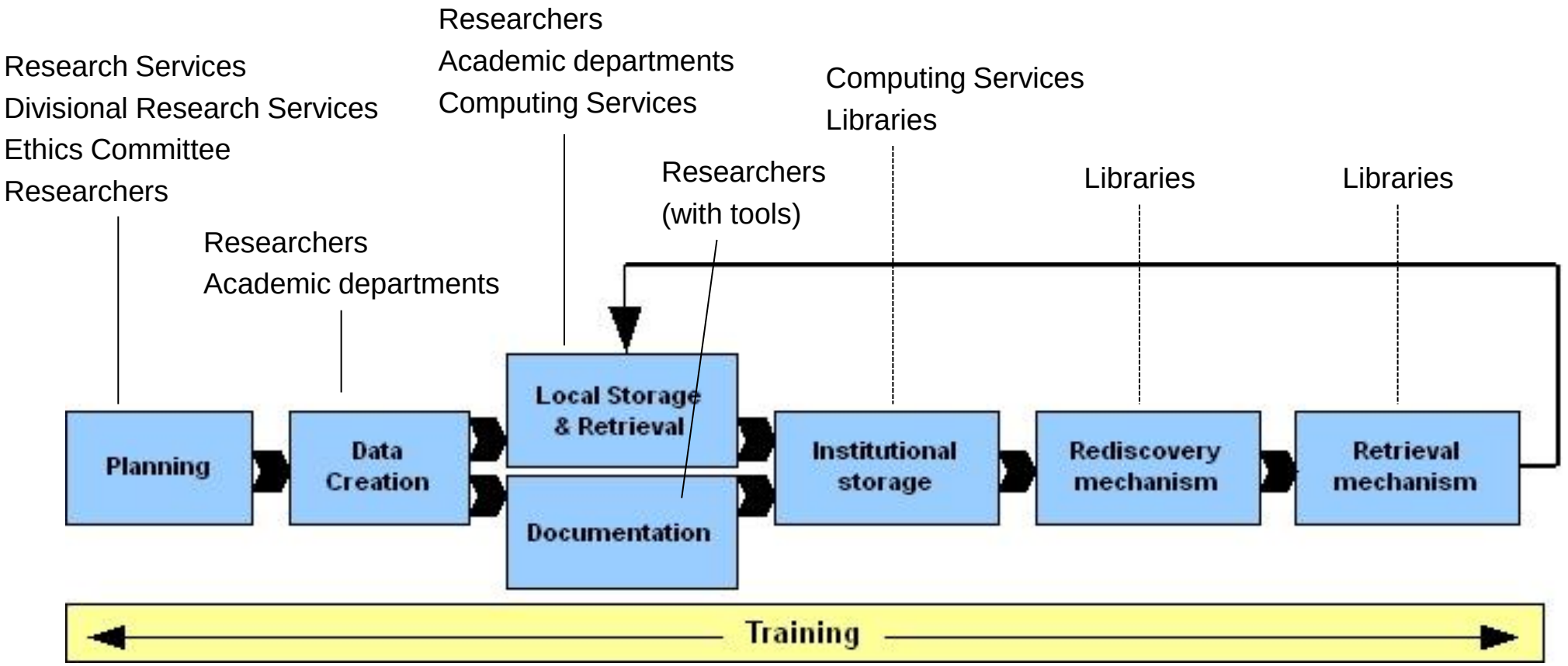
JISC

UNIVERSITY OF OXFORD

Institutional structure

- University of Oxford has a highly federated structure
  - Principal of subsidiarity
- Developing a data management infrastructure is not something that one part of institution can undertake alone
- Computing Services taking the coordinating role
  - Already maintained some infrastructure
  - Tradition of working closely with researchers
  - Office of director of IT embedded within department

Elements of data management infrastructure



Research Services; Divisional Research Services; academic divisions; academic faculties; Computing Services; Libraries.

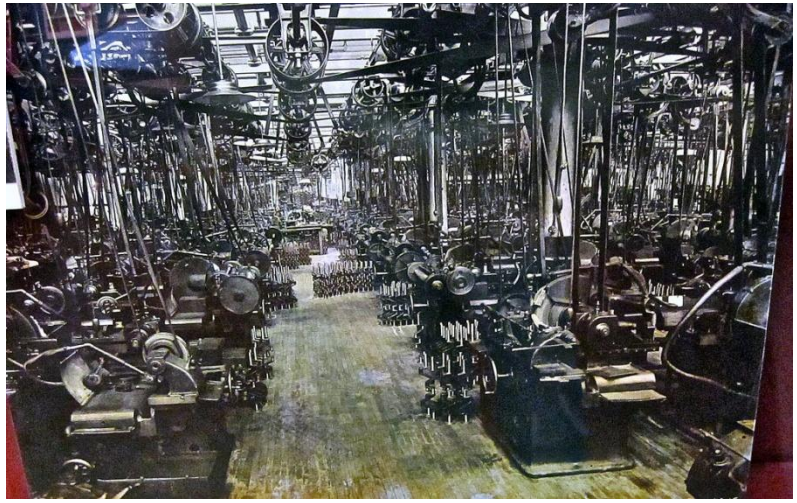
Why bother with institutional data preservation?

- Not all academic disciplines covered by national data centres, nor ever likely to be
- Assistance as grant proposal stage
- Reputation management
- Sustainability
- Is it cost effective?
  - Data centres can offer cross-institutional economies of scale
  - Data centres can act as hubs of specialized expertise

High and Low curation

- High Curation
  - High levels of expertise
  - Human intervention at ingest
  - Metadata cleaning and standardisation
  - High levels of long-term care & curation

- Low Curation
  - Largely automated
  - Little quality control at ingest
  - Generic service



JISC

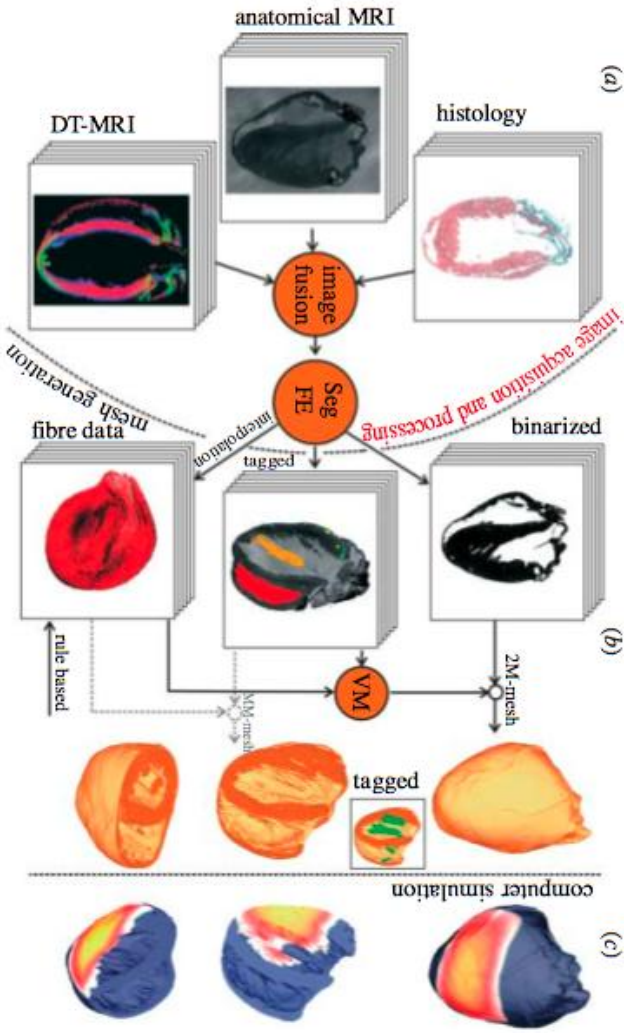
UNIVERSITY OF OXFORD

Approach

- 2006 – formation of Oxford Digital Repositories Steering Group (ODRSG)
- 2008 – internally-funded project ‘Scoping Digital Repository Services for Research Data Management’
  - Established researcher requirements; evaluated current service provision
- 2009 – two JISC-funded projects
  - Embedding Data Curation Services in Research (EIDCSR)
  - Supporting Data Management Infrastructure for the Humanities (Sudamih)
- Attack on all fronts!
  - Don’t lose sight of the interrelated nature of data management activities
  - Get all service departments involved

EIDCSR

- Working with 3D Heart Imaging Project
- User requirements
  - Secure storage (5 year mandate)
  - Tools for rapidly visualising (and sharing) data
  - Metadata (for rediscovery of data)
- Data outputs (approx 1TB per heart)
  - Histology (very large 2-D images)
  - Anatomical MRI (3-D images)
  - Diffusion Tensor MRI (3-D images)
  - Segmentation & Mesh data
- + Institutional Data Management Policy

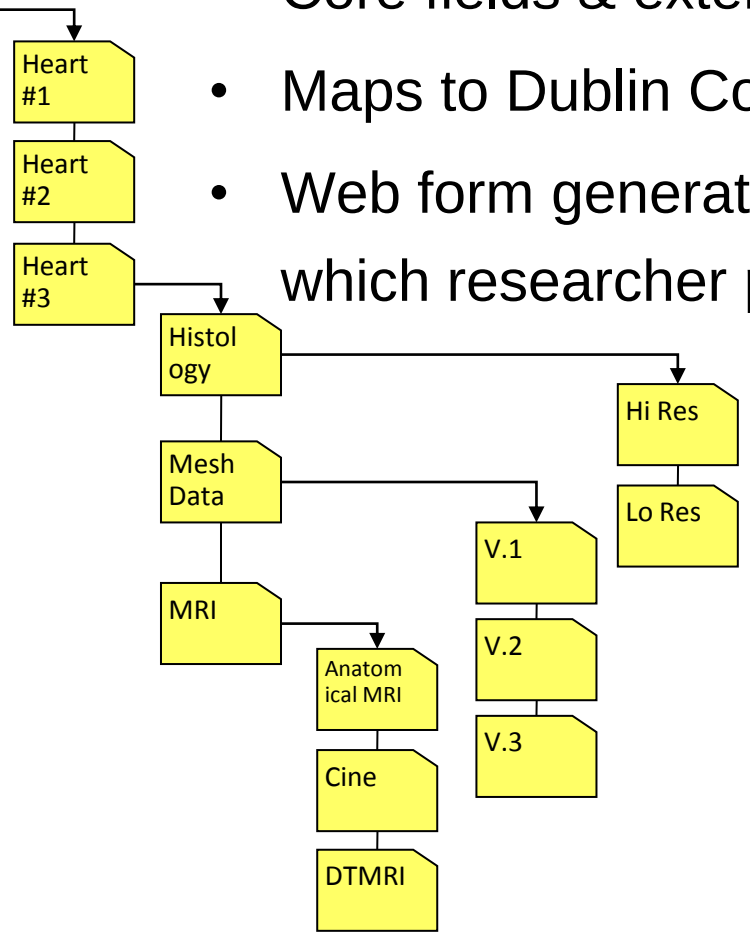


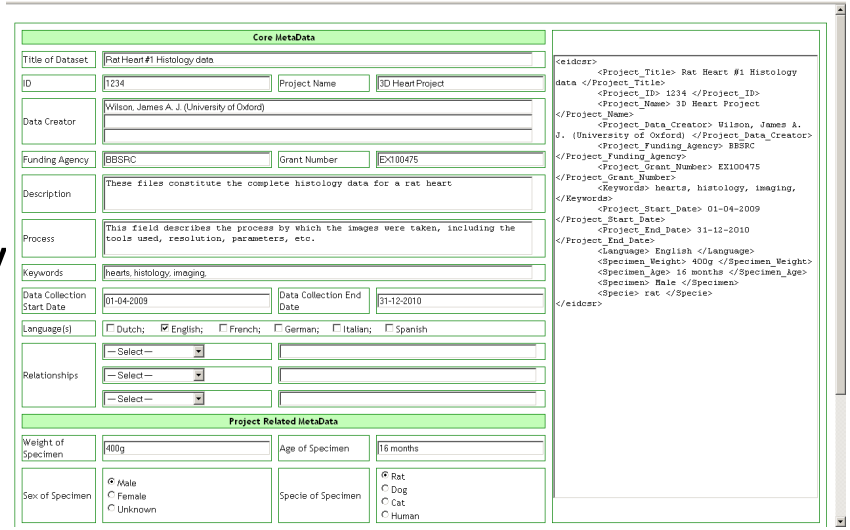
JISC

UNIVERSITY OF OXFORD

Data & Metadata

- Core fields & extensible user-defined
- Maps to Dublin Core (mostly)
- Web form generates XML read-me files which researcher places in data directory



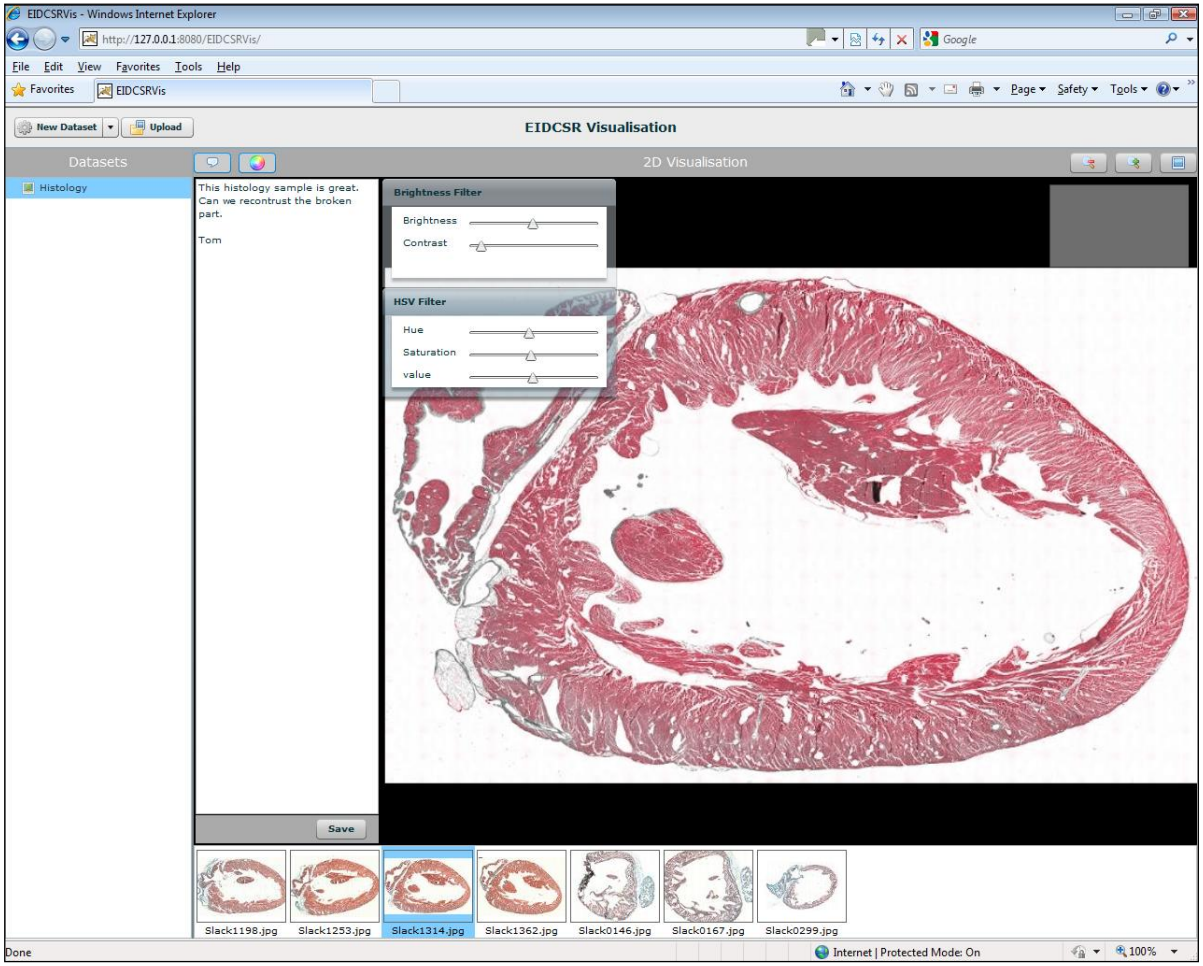


- File structure archived to HFS
- Metadata interpreted during archive process & sent to Libraries' 'Databank' system
- Re-archiving updates metadata

JISC

UNIVERSITY OF OXFORD

Visualisation software - Workbench



- Enables 2D and 3D views
- Web interface
- Zooms to enable very high definition viewing
- Images may be annotated
- Different permissions levels
- Thresholding tools

JISC

UNIVERSITY OF OXFORD

Institutional Policy

- Part of a wider programme of research integrity led by the Research Services Office
  - Consultation with University of Melbourne
- Must involve researchers
- Requires top-down and bottom-up approaches
  - Departments need to interpret central policy and produce local versions
- Requires awareness raising
- Requires development of support service to actually implement recommendation
- Slow process!

Sudamih

- Better understanding of research data management practices and needs in the humanities
- Development of training modules to improve information/data management skills
- Development of a ‘Database as a Service’ (DaaS) system
- Cost models for data curation services

JISC

UNIVERSITY OF OXFORD

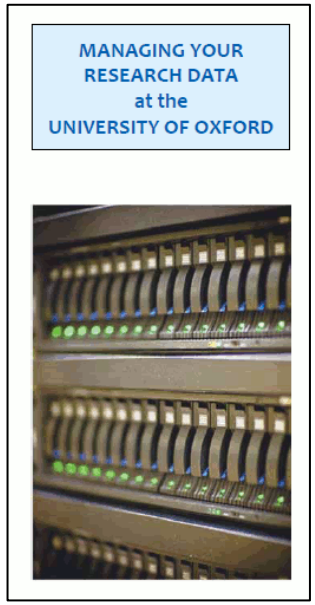
Training Requirements

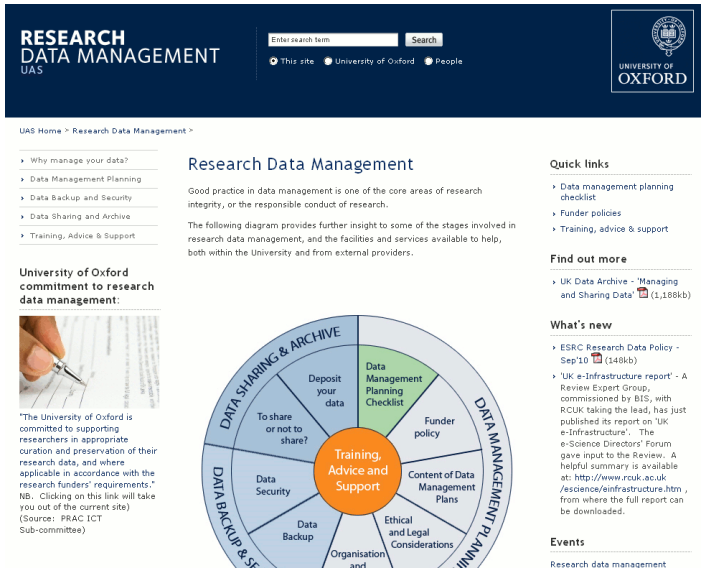
- Researchers not familiar with concept of ‘data management’
- Not directly addressed by existing training
- Areas of concern to researchers:
  - organizing one's files
  - linking notes to content
  - keeping track of sources / data integrity
  - backing up; versioning
  - software tools for particular research challenges
  - structuring data in databases
- Also demand for a technical consultancy service

“Most people are so inundated with opportunities to attend training and conferences and workshops that they don’t have time to take up many of them. People tend not to worry about data management until it becomes an issue and there’s something specific they need to do” [Music Faculty Lecturer]

Training Plan

- Training does not fall neatly under one service group's remit
- Researchers not especially interested in data management
- Therefore, try to integrate training into existing infrastructure





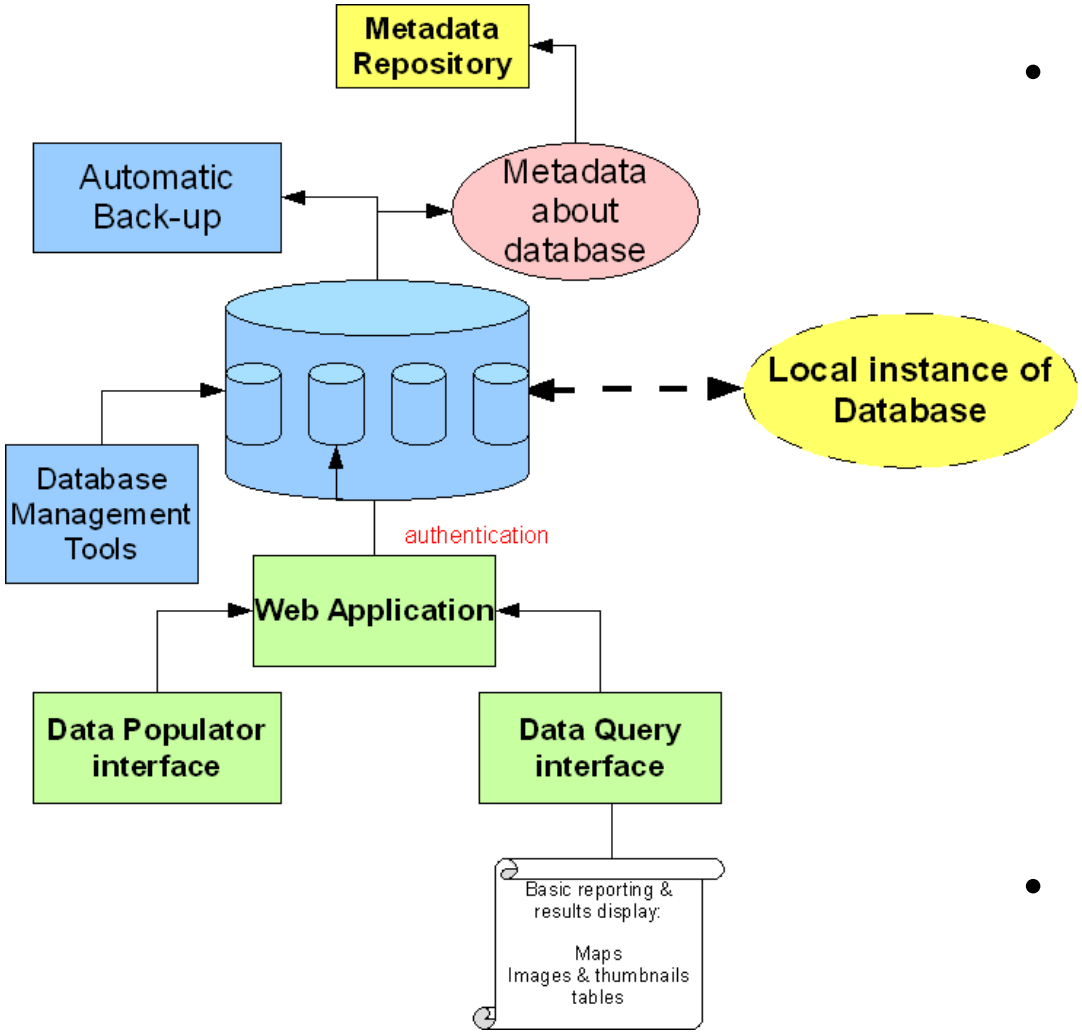
Five priority areas:

1. Introduction & existing services
2. Tools to help manage data
3. Organise & link information
4. Technical aspects of funding bids
5. Database design for the humanities

JISC

UNIVERSITY OF OXFORD

Database as a Service (DaaS)



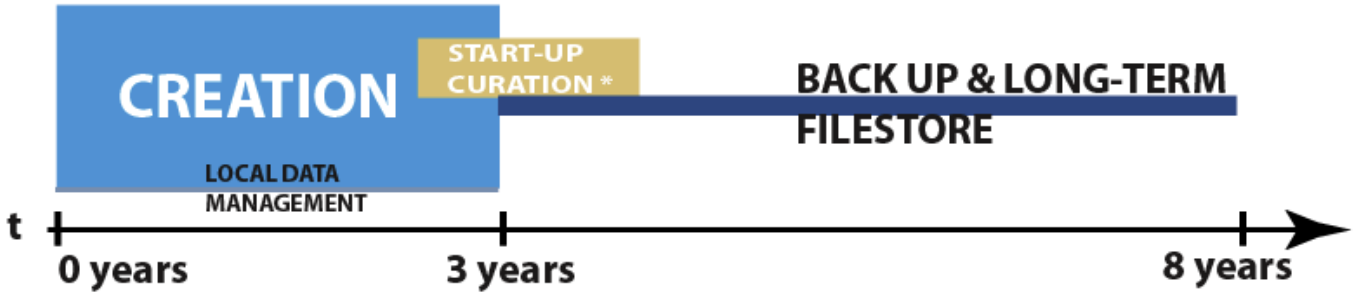
- Web-based system for creating simple relational databases
  - centrally hosted and maintained
  - regularly backed-up
  - easily shared with collaborators and the public
  - capable of dealing with text, images, and geospatial data (initially)
  - Described and discoverable
- Helps with data creation, storage, and documentation

JISC

UNIVERSITY OF OXFORD

Cost models & business cases

- Assistance from JISC Managing Research Data programme
- Costs relatively easy to assess; benefits less so
  - Building on existing infrastructure helps
  - Basing costing upon a Service Level Description
- Contributed to Keeping Research Data Safe projects (KRDS)



- Low-curation approach to keep costs low – responsibility concentrated with researchers, rather than ‘curators’

Lessons learnt

- Mind your language
- Different elements work at different speeds/rhythms
- Communication and engagement are key. Need buy-in at multiple levels
- Make the most of existing infrastructure
- Trade-off between comprehensiveness and usability
- Researchers are key, but they probably don’t think all that much about re-usability, long-term curation, etc.

Thanks!