

Disseminating Oxford's research in the
era of electronic communications



Research Data Management – Can we do it?

Paul Jeffreys, Director of IT

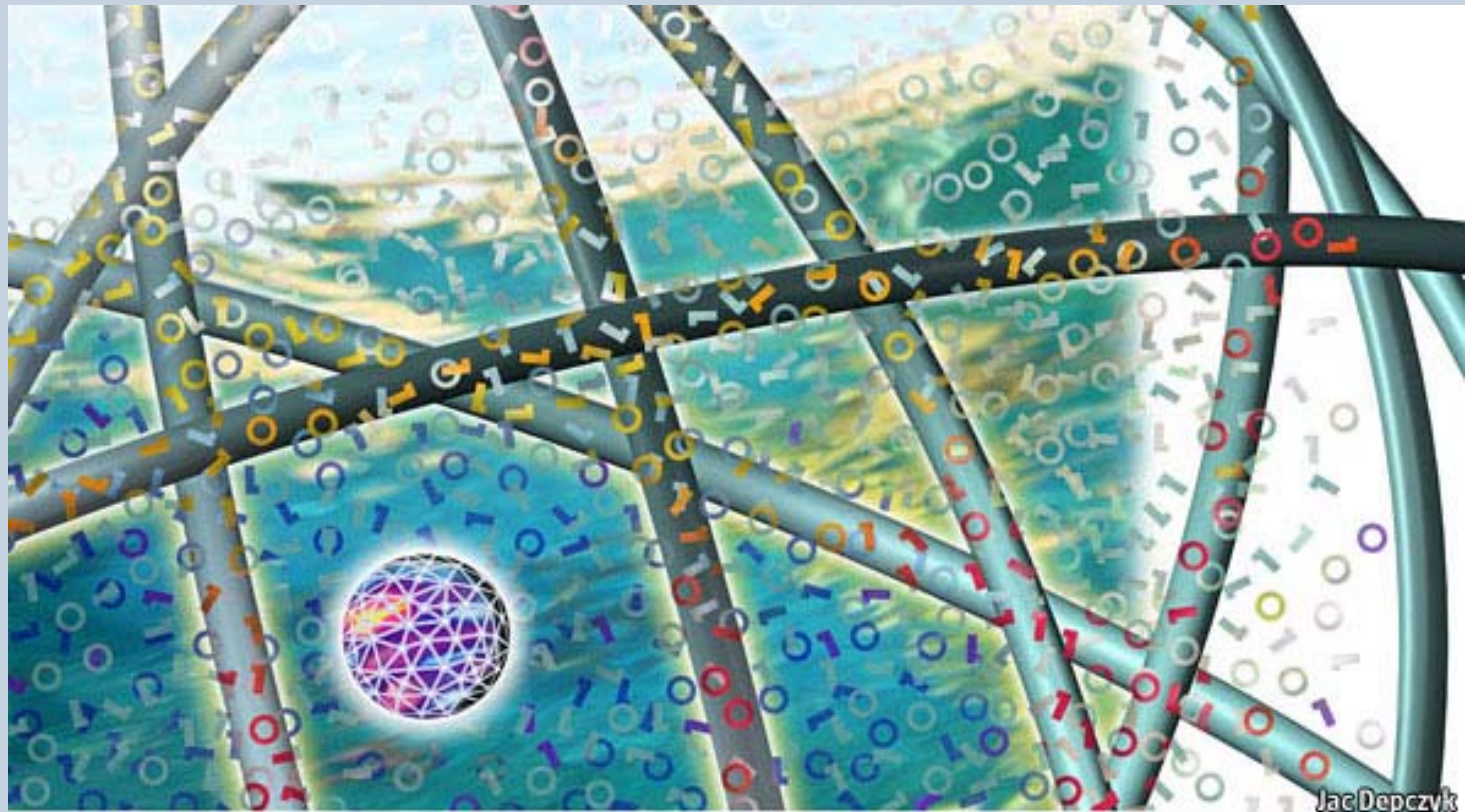
EPA Lecture Theatre, 4 March 2010

Economist: A special report on *Managing Information*

Data, data everywhere

Information has gone from scarce to superabundant. That brings huge new benefits, says Kenneth Cukier - but also big headaches

Feb 25th 2010



Paul Jeffreys, *Research Data Management*
Disseminating Oxford's research in the era of electronic comms.

Economist Article (1)

The data deluge

Businesses, governments and society are only starting to tap its vast potential



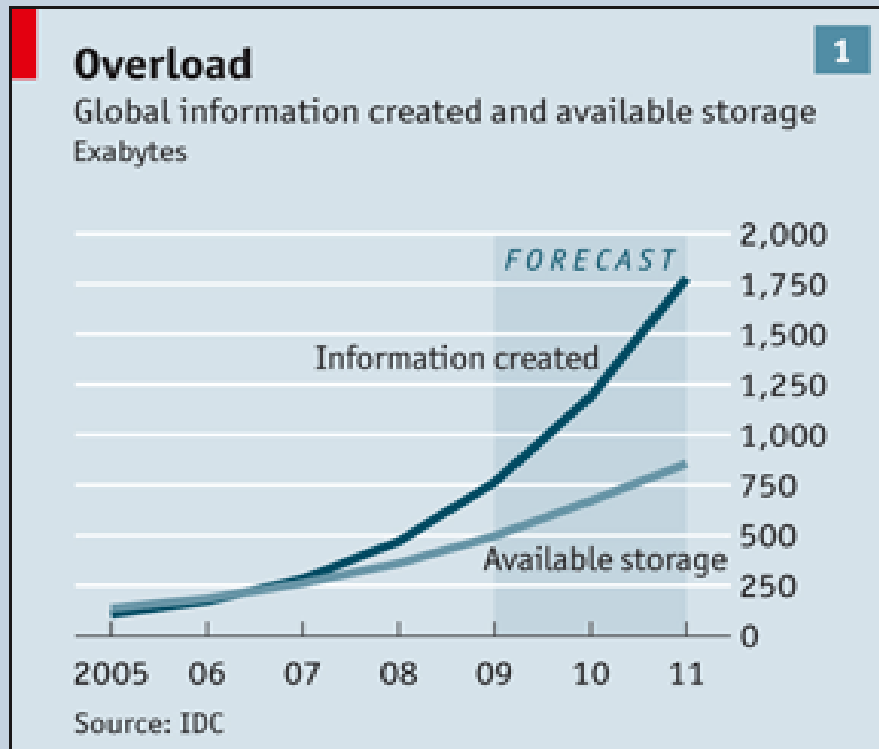
‘Everywhere you look, the quantity of information in the world is soaring. According to one estimate, mankind created 150 exabytes (billion gigabytes) of data in 2005. This year, it will create 1,200 exabytes.’

‘Joe Hellerstein, a computer scientist at the University of California in Berkeley, calls it “the industrial revolution of data”.’

the business of information management—helping organisations to make sense of their proliferating data—is growing by leaps and bounds. In recent years Oracle, IBM, Microsoft and SAP between them have spent more than \$15 billion ...’

Economist Article (2)

‘Given enough raw data, today’s algorithms and powerful computers can reveal new insights that would previously have remained hidden.’



Chief information officers (CIOs) have become somewhat more prominent in the executive suite, and a new kind of professional has emerged, the data scientist, who combines the skills of software programmer, statistician and storyteller/artist to extract the nuggets of gold hidden under mountains of data.’

‘The data-centred economy is just nascent,’ admits Mr Mundie of Microsoft. “You can see the outlines of it, but the technical, infrastructural and even business-model implications are not well understood right now.”

Economist Article (3)

Towards the end of the article:

‘Needle in a haystack’

‘As data becomes more abundant, the main problem is no longer finding the information [initially] as such but laying one’s hands on the relevant bits easily and quickly’

‘New rules for big data’

‘Ensuring the integrity of information is an important part of the big-data age’

‘Handling the cornucopia’

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention”, Herbert Simon, an economist, 1971’

Research Data Management (1)

- A subset of *Managing Information*
 - Although many aspects of the Economist's report are relevant
- Example: doctorate theses
- Research data in universities:-
 - Create by researchers
 - Access data created by government/industry/public sector/heritage
 - Store – in some form
 - Refine, visualise, abstract and store
 - Annotate
 - Augment
 - Use to derive publications
- Approaches very different across disciplines

Research Data Management (2)

- Often - would like ability to:
 - Search, locate (re-locate) and reload some *selection*
 - Aggregate some *selections*
 - Store *selection* in suitable form that capable of re-use
 - Manage *selection* within 'life's work' context
 - Long term preserve *selection*
 - Deposit some *selection* in disciplinary, national and University archives
 - Publish some *selection*, and make direct connection with journal publications
- Meet data policies specified by funding agencies
- Enable researchers to manage with negligible overhead

Rest of talk will follow two entwined strands:

Oxford's work in this area:

- Oxford Digital Repositories Steering Group
- Scoping and Developing Institutional Data Services
- JISC projects:
 - EIDCSR
 - SUDAMIH
 - ADMIRAL
- University Context
- Stock Take

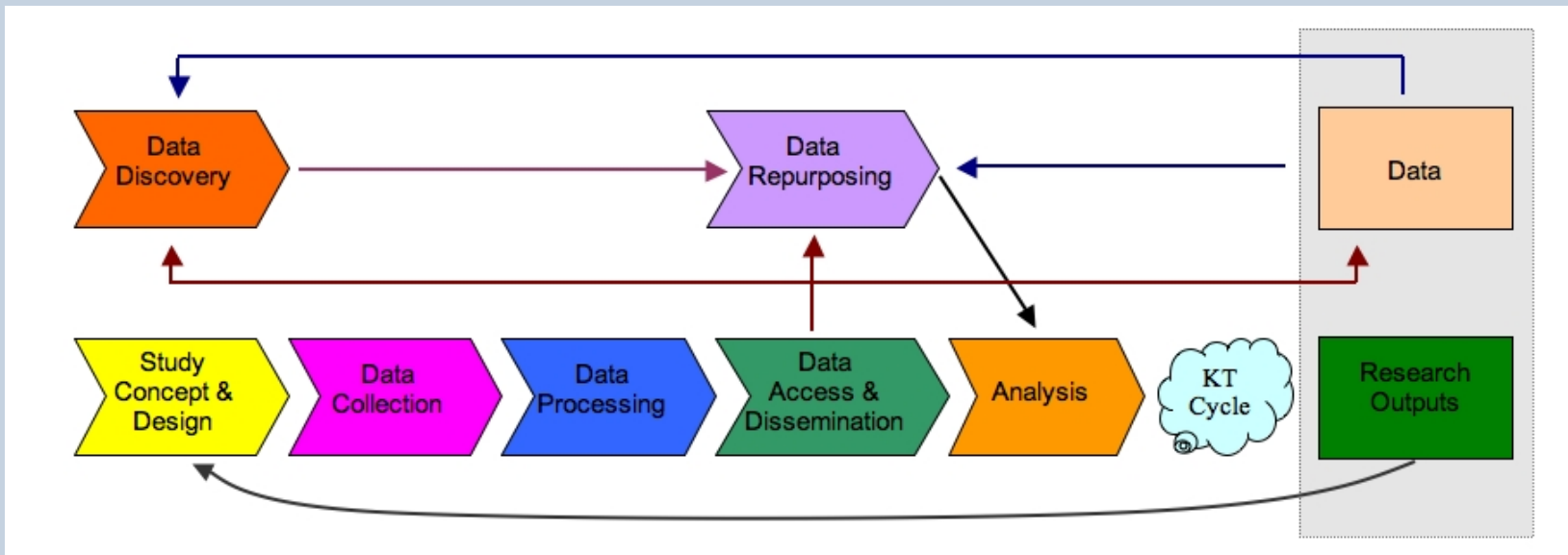
Inter-National activity:

- Research Life Cycle
- Digital Curation Centre (DCC)
- DCC Curation Lifecycle Model
- UK Research Data Service (UKRDS)
- Data Audit Framework
- Research e-Infrastructure

Chuck Humphrey's Research Cycle

<http://datalib.library.ualberta.ca/~humphrey/lifecycle-science060308.doc>

Bottom set of links is research cycle, relating to data; incorporates a knowledge transfer cycle above

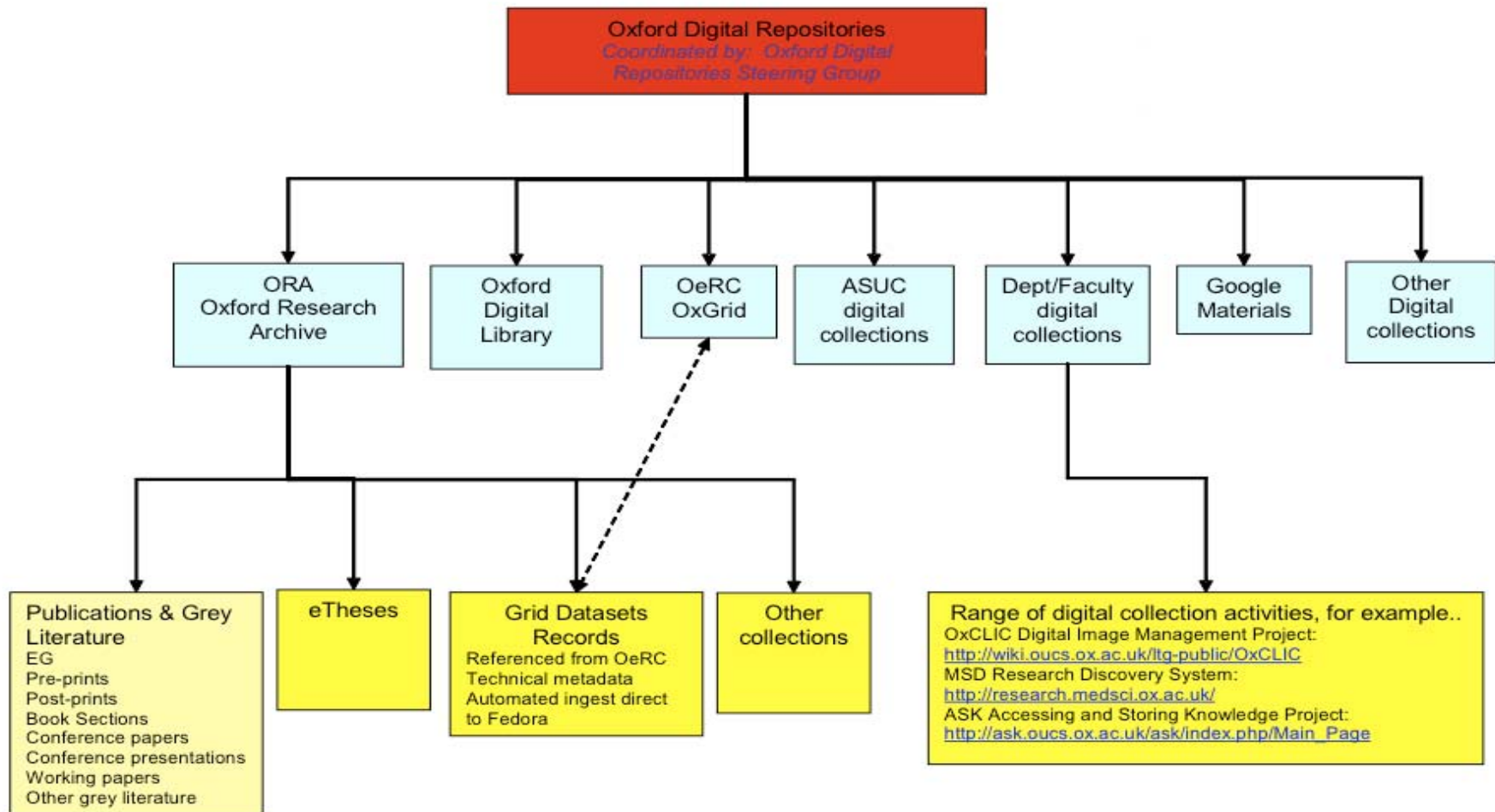


Digital Curation Centre

<http://www.dcc.ac.uk/>

- “The Digital Curation Centre is the UK’s leading centre of expertise in digital data curation
- Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle
- The active management of research data reduces threats to their long-term research value and mitigates the risk of digital obsolescence. Meanwhile, curated data in trusted digital repositories may be shared among the wider UK research community
- As well as reducing duplication of effort in research data creation, curation enhances the long-term value of existing data by making it available for further high quality research”

Oxford Digital Repositories Steering Group



Scoping Digital Repository Services for Research Data Management

http://www.ict.ox.ac.uk/repositories/index.xml.ID=body.1_div.2

OUP John Fell Fund +
JISC funded DataShare
UKRDS collaboration



From Jan'08 to March'09

Aim: to scope the requirements for a digital repository service to store, disseminate and preserve research data generated at Oxford

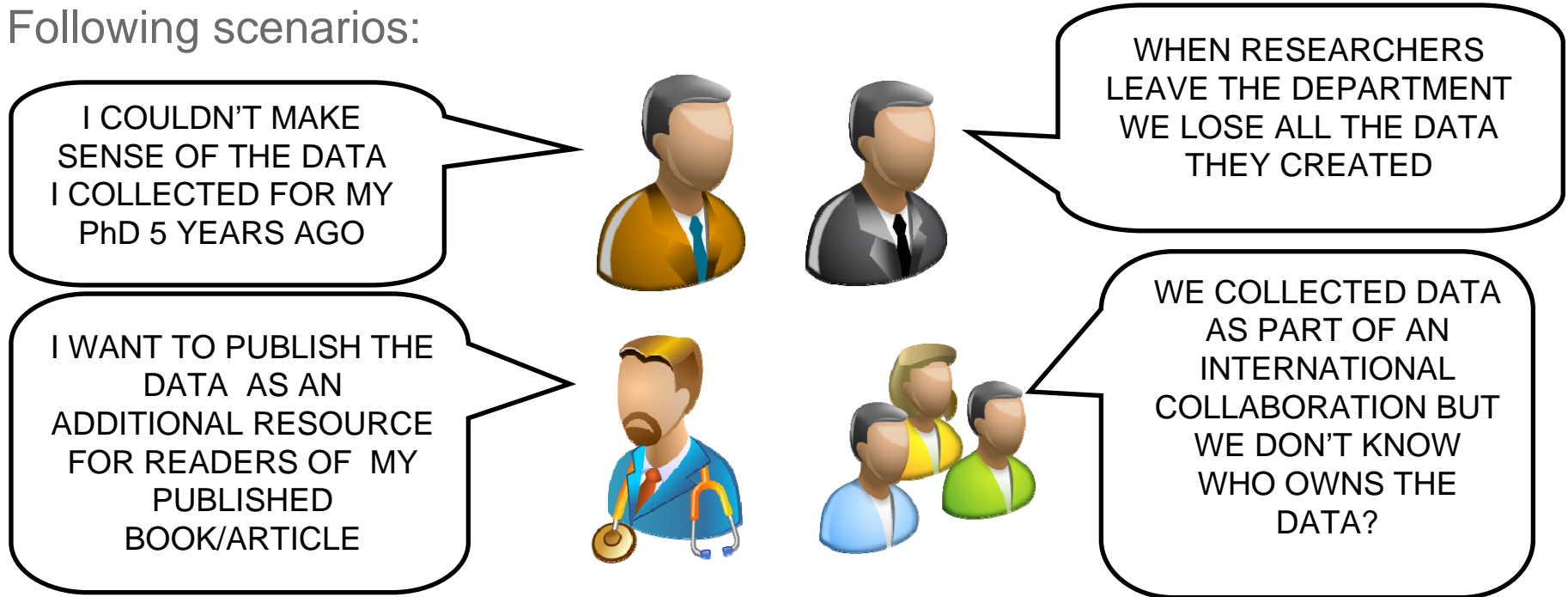
Interviews with researchers

Around 40 interviews

Part of the Oxford case study work for UKRDS feasibility study



Following scenarios:



Deliverables from Scoping Services survey

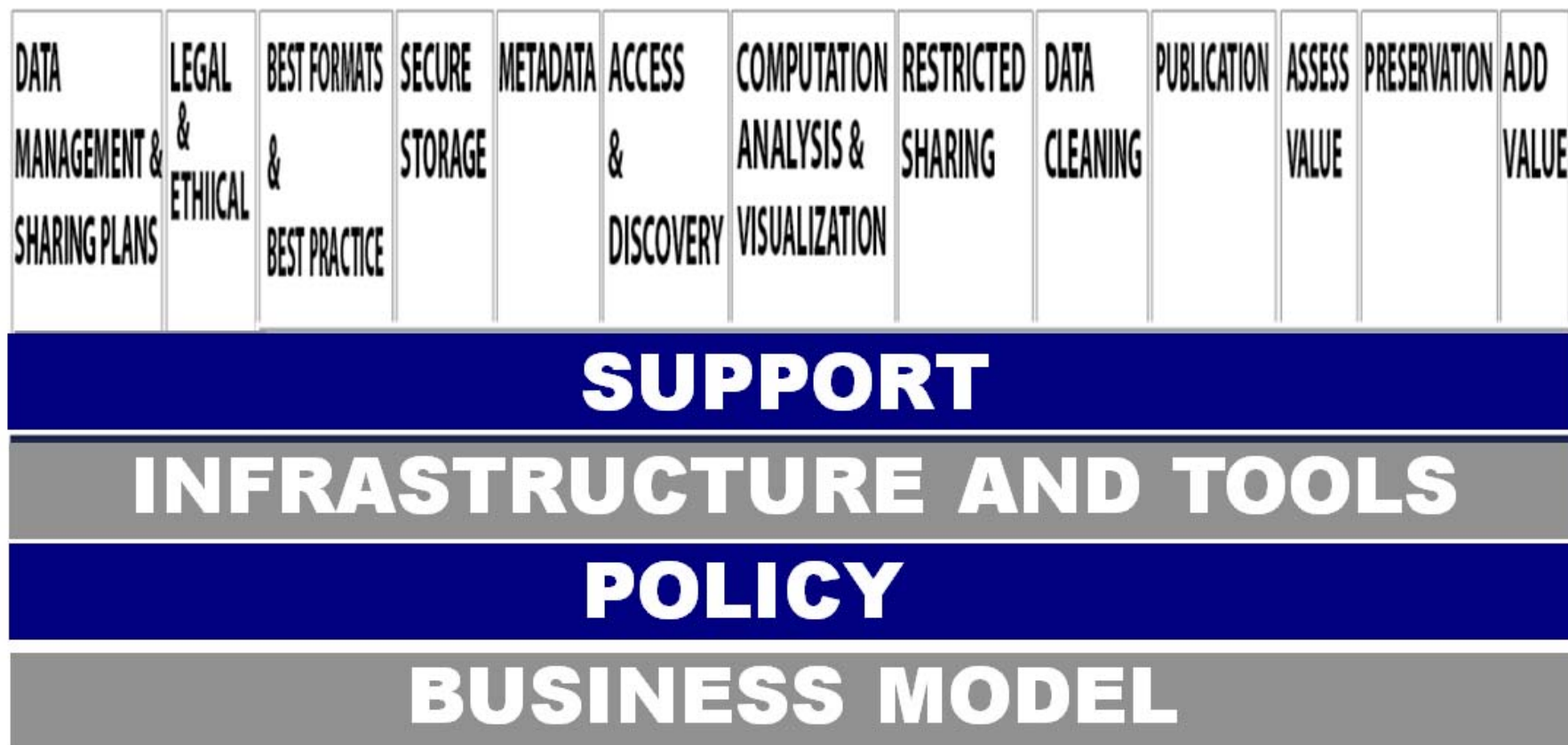
Researcher Requirements:

- Advice and support across research lifecycle
- Secure storage for large datasets with access controls
- Sustainable infrastructure for publication and long-term preservation

Service Unit Findings:

- Widespread expertise in data management and curation amongst service units in Oxford
- Support provided in ad-hoc basis, services not made explicit
- Overall, majority of the services in the data management and curation framework are not offered fully or at all
- A need for a university wide policy on data management and curation

Research data management and curation services framework



UK Research Data Service (UKRDS)

<http://www.ukrds.ac.uk/>

- Joint project between RLUK (a consortium of research libraries), and RUGIT
- Funded by HEFCE under its Shared Services programme, with support from JISC
- Initial objective: assessing feasibility and costs of developing and maintaining a national shared digital research data service for UK Higher Education sector
- Seen by project sponsors: forming crucial component of the UK's e-infrastructure for research and innovation, and one which will add significantly to the UK's global competitiveness
- Feasibility study concluded:
 - embedding skills, capability and organisation into the HEI research management process was the best approach
 - relatively small national service structure would be needed to foster this through channelling training, tools and good practice developed by existing national and international skill centres

Embedding Institutional Data Curation Services in Research (EIDCSR)

<http://eidcsr.oucs.ox.ac.uk/>

Funded by 

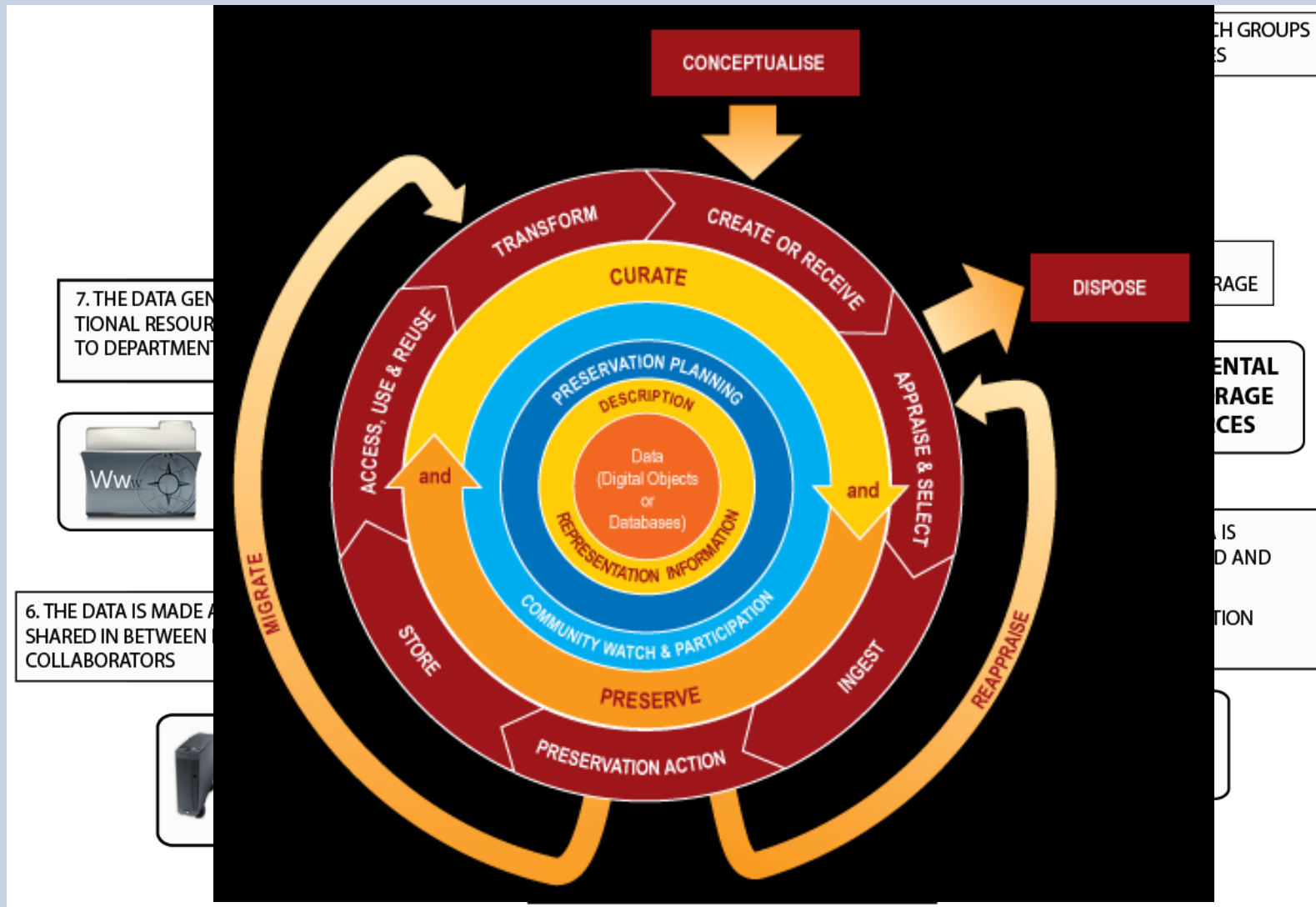
From April 2009 to September 2010



Cardio Mechano-Electric Feedback Group
Computational Biology Group

Aim: to address the data management and curation requirements of two collaborating medical research groups in Oxford

EIDCSR Curation Lifecycle



The Data Audit Framework (DAF)

- <http://www.data-audit.eu/>
- DAF provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets
- Developed in University of Glasgow, JISC funded
- EIDCSR used DAF methodology to:
 - document research group's data management practices
 - map research data resources within the two research groups and
 - compile information about data resources freely available on the Oxford University web pages



Supporting Data Management Infrastructure for the Humanities (SUDAMIH)

<http://sudamih.oucs.ox.ac.uk/>

Funded by



From October 2009 to March 2011



**Humanities Division Research Support Team
Faculty of Classics (Archaeology)
Faculty of History**


Aim: to address a coherent range of requirements for the more effective management of data (broadly defined) within the Humanities at an institutional level

SUDAMIH – Project Overview

- **Effective management of data within the Humanities at an institutional level**
 - -> with methodologies, outputs, outcomes useful generically
- **Overarching "life's work" theme**
 - online database service to enable researchers to create online databases with support of textual, image and geo-data
 - data management support and training for researchers at various stages including personal data management, creation of online databases and management of hybrid data collections
- **Utilise same framework developed for EIDCSR**
 - Multi-service approach
 - Scope research requirements
 - Implement infrastructure tools and support
 - Develop costing and sustainability models
 - Develop policy frameworks

A Data Management Infrastructure for Research Across the Life sciences (ADMIRAL)

http://imageweb.zoo.ox.ac.uk/wiki/index.php/ADMIRAL#The_ADMIRAL_Project:_A_Data_Management_Infrastructure_for_Research_Across_the_Life_sciences

Funded by 

From October 2009 to March 2011

Image Bioinformatics Research Group
Oxford University Library Services
British Library
MIT

Aim: to create a two-tier federated data management infrastructure for use by life science researchers, to provide services (a) to meet their local data management needs for the collection, digital organization, metadata annotation and controlled sharing of biological datasets; and (b) to provide an easy and secure route for archiving annotated datasets to an institutional repository, The Oxford University Data Store, for long-term preservation and access, complete with assigned Digital Object Identifiers and Creative Commons open access licences.

University of Oxford Context

Strategy and Policy

**Research
Committee**

PRAC ICT

RIMSC

Researcher and Services

**Academic
Divisions**

**Research
Services**

**Library
Services**

**Computing
Services**

Strategic projects

**Scoping
Digital
Repository
Services**

ADMIRAL

EIDCSR

SUDAMIH

UK RDS

University Statement

“

The University of Oxford is committed to supporting researchers in **appropriate curation and preservation of their research data**, and where applicable in accordance with the research funders' requirements. It recognises that this must be achieved **through the deployment of a federated institutional data repository**. This repository has to be supported by a suitable business model, and where possible funded through full economic cost recovery, so that the University can guarantee that the data deposited there will be managed over the long term. The data repository will be a cross-agency activity **developed and supported by a number of departments within the University and will build, as far as possible, on existing services**, including the Oxford University Research Archive (ORA). It will be **overseen by a Steering Group which reports to the University Research Committee**. The management and curation of research data will be addressed **in cooperation with specialist agencies**, research funders and other institutions in the UK and internationally. Oxford is committed to playing a significant role within the foreseen UK Research Data Service Pathfinder activities”

UKRDS update

- Interim planning phase of project funded to:
 - Design and develop service structure
 - Plan a Pathfinder implementation with case study HEIs: Bristol, Leicester and Oxford
 - Plan national 'registry'
 - Build of overall strategy to 'embed capabilities in HEIs'
- New Director of DCC just announced
- Generally recognised that Oxford is significantly ahead of other HEIs
 - In essence three/four projects ahead...
- Despite all the financial uncertainty, UKRDS encouraged to make a bid to HEFCE before May 2010
 - Very significant component relates to Oxford

Related National Developments

1. RCUK Review of e-Science:

<http://www.rcuk.ac.uk/escience/townmeet.htm>

“Data stewardship at enormous scale

Continue the strong focus on creating practices and services for appraisal, curation, federation, and long-term access to scientific data. Complement or broaden the activities of the Digital Curation Centre with the creation of coordinated and sustained production services for curation and stewardship of scientific data. Consider a highly centralised, large data centre model for storing the information and preserving the bits, together with a distributed model for curation by disciplinary specialists. The academic digital libraries centres might be encouraged to assume some major responsibility for scientific data. Seek and promote International cooperation. In all of this, plan for continued exponential growth in scientific data.”

2. OSI’s report – ‘Developing the UK’s e-Infrastructure for science and innovation’:

<http://www.rcuk.ac.uk/escience/news/einfrastructure.htm>

- Under revision by Professor Carole Goble and team, building on RCUK Review
- Fundamental importance for future national research infrastructure
- Will report in next month

Stock Take: Can we manage our research data?

- Superb foundation laid by Oxford Digital Repository Coordinator
- Three JISC-funded projects taking us forward in specific disciplinary areas and giving us particular capabilities
 - Must be emphasised that very small part of whole...
- RIMSC crucially important for overseeing and owning the activity
- Take-up of research data management services, and types of service needed will inevitably vary significantly across the disciplines
- Three projects will end by spring 2011 – possibility of further JISC funding?
- But clear- the University will need a strategy and will need coordinated research data management
 - Not least to satisfy the research funding organisations
- If the UKRDS pathfinder service is funded, then Oxford will be able to continue to develop its research management activities using external funds
 - Emphasis: sustainable - leadership, infrastructure and training
 - And, develop to the point that can operate on *cost recovery basis*

Conclusions

- Research data management must be developed for the University of Oxford in close collaboration with researchers
- Will be required to satisfy those who fund research
- Make our research more efficient and effective
- The types of service, functionality, relevance, take-up will vary across disciplines
- Oxford is one of the leading institutes in the UK
- Exciting projects are underway
- All activities have been externally funded thus far
- If UKRDS is funded by HEFCE, further external funding
- High level commitment and underwriting from the University will be essential

Acknowledgement and thanks, in particular, to: Mike Fraser, Luis Martinez, James Wilson, David Shotton, Graham Klyne, Sally Rumsey, Erin Cooper, Kathryn Dally, Simon Hodson (JISC), John Milner (UKRDS)