Institutional Data Management Blueprint

Kenji Takeda (Engineering Sciences), Mark Brown (University Librarian), Simon Coles (Chemistry), Les Carr (ECS, Eprints), Jeremy Frey (Chemistry), Peter Hancock (iSolutions), Graeme Earl (Archaeology)
Introduction

- Institutional Data Management Blueprint JISC project
- Latest results
- Pilot studies
- Conclusions

www.southamptondata.org
Data Management @ Southampton

- What do we mean?
  - Everything
- Why do we care?
  - Foundation for all of our research
- How should it be managed?
  - We want to find out from you
- How can the University help?
  - What do you need?
IDMB Project Overview

- Produce framework for managing research data for an HEI
  - based on an analysis of current data management requirements for a representative group of disciplines with a range of different data.
  - Building on the developed policy and service-oriented computing framework

- Scope and evaluate a pilot implementation plan for an institution-wide data model
  - integrated into existing research workflows, and
  - extend the potential of existing data storage systems, including those linked to discipline and national shared service initiatives.
Institutional Data Management: Curation & Transformation

COMMUNITY
- Journals
- Blogs
- Forums

SEMANTIC AUTHORING
- Workflow & Analysis Services
- Access & Preservation Repository

SEAMLESS INGEST TOOLS
- Experiment & Measurement
- Observation & Survey
- Calculation & Simulation

INSTITUTION
- Research-led Learning & Teaching

COLLABORATION
- Trust & Evaluation Services

INFORMATION
- EdShare

DATA
- eprints
- Crystals
- myexperiment
Project Team

• Project team:
  – Kenji Takeda (Engineering Sciences, WP leader), Mark Brown (University Librarian), Simon Coles (Chemistry), Les Carr (ECS), Jeremy Frey (Chemistry), Peter Hancock (iSolutions), Graeme Earl (Archaeology), Steve Patterson (iSolutions), Wendy White, Fiona Nichols, Michael Whitton, Harry Gibbs, Christine Fowler, Pam Wake (Library)

• Steering Group:
  – Philip Nelson (DVC Research), Adam Wheeler (Senior DVC, Infrastructure/Planning), Jeremy Frey (Chemistry), Helen Snaith (National Oceanography Centre), Simon Cox (Engineering Sciences, EPSRC HPC Technology Watch Panel), Graham Pryor (DCC), Sally Rumsey (University of Oxford), Ian Tebbett (iSolutions, Director), Kenji Takeda (Engineering Sciences), Mark Brown (University Librarian)
Researcher-centric

• Researcher focussed project
• Want to know how researchers currently manage, store and archive your data
• Want to know how this can be improved
• Only then can the University plan effectively

www.southamptondata.org
Three-legged stool data model
Data Management Development

- Assess where we are (inc. AIDA)
- Discover what we do well
- Find out where the gaps are
- Help to plan what needs developing
- Raise awareness
- Develop best practice
- Training and advice
Work Packages

- **WP1**: Policy development and implementation
  - Infrastructure, governance, cost-benefit, roadmap
- **WP2**: Best practice
  - Audit, workshop, gap analysis, best practice implementation strategy
- **WP3**: Pilot implementation
  - Archaeology, nano-fabrication, discipline cluster meta-repository
- **WP4**: Dissemination
  - Web, workshops, postgraduate training course
Questionnaire & interviews

- Archaeology, Chemistry, Electronics & Computer Science, Engineering Sciences
  - Roll-out to University pending ethics clearance
- Online Questionnaire
- Follow-up interviews
- Suggestions box and polls @
  - www.southamptondata.org
Feedback so far

- Who is responsible for managing your data?
Feedback so far

- Where do you store your data?
Feedback so far

• How much electronic data do you currently retain?
Feedback so far

• How long do you keep your data for?
Feedback so far

• How frequently do you backup your data?
Feedback so far

• Where do you backup your data?
Feedback so far

• How do you keep track of where your data is stored and relates to?
Feedback so far

- Would you find it useful to have university-wide guidelines to manage and maintain your research data?
Discussion @ Kick-off workshop

Quick wins

Dreams

Issues
Quick Wins

- Ask Tesco/BP how they do it!
- University standards / protocol guidelines.
- Data management pro-forma.
- Seminars about data management.
- Advice on data labelling.
- Make Graduate School data management course compulsory.

- Are people aware of existing data services at the university?
- No quotas for file storage.
- Secure data more quickly and risk analysis of data loss.
- Stop single point failure of hard drives.
- Easy access to repositories across departments.
- Repository for data management plans.
Dreams & Aspirations

- Visibility of my research history and related colleagues / students.
- Seamless integration from papers to source data.
- Central data body linking all university / RC archives together.
- Intuitive and natural system, making putting data in and getting it out “child’s play”.
- Automated archive and meta-data generation.
- Knowledge base for sharing research data and resources – including bibliography lists, external data sets, contacts etc.
- Expert support for planning and implementing data management plans Kudos for good data management.
- Integrated policy approach by RCs internationally.
- Capturing data in electronic form, without overhead (e-lab note books).
Issues & frustrations

• Meta data: how to make it meaningful and useful later is hard to define upfront.

• Ongoing curation costs and implications are not fully understood.

• Responsibility of data generated by former university members.

• Researcher’s exclusive use of own data: guidelines, times etc.

• Viewing / reprocessing data if you do not have the original program installed.

• Multiple institution / university guidelines.

• Legal issues; FOI requests indicate that the public “owns” the data, not the university.

• Security: Who has access and where? What are the implications of a compromise? Who is responsible?

• Should we curate bad data?

• Lost data.
Southampton Nanofabrication Centre

• Help nanoscience researchers in new Mountbatten Building
  – Helium Ion microscope
  – ASM Epsilon Epitaxy System
Conclusions
Expected Outcomes

- Pathfinder for institutional data management strategy for next decade;
- Data management institutional blueprint based on an analysis of data management requirements and current best practice;
- Service-oriented (including RESTful), extensible enterprise architecture model for data management;
- 10-year business model roadmap;
- Best practice gap analysis report;
- Pilot implementation for infrastructure, human and technological;
- Workshops, training, website and reports for dissemination of best practice.
What’s in it for researchers?

• Guidelines and best practice
  – Advice and support from the University

• Training for you and your researchers
  – Graduate School modules

• Better University and School systems designed for you
  – Supported by iSolutions and the Library

• An easier life!
  – Enabling you to concentrate on your research
Conclusions

• Good data management is vital for better research

• Two-pronged approach
  – Bottom-up to augment researcher’s world
  – Top-down to provide support and guidance

• www.southamptondata.org

• ktakeda@soton.ac.uk