The benefits of more effective research data management in UK Universities
Manjula Patel & Neil Beagrie
UKOLN, University of Bath & Charles Beagrie Ltd.

JISC Conference 2011: Financial Challenges – Digital Opportunities
BT Convention Centre, Liverpool
Tuesday 15th March 2011
I2S2 Project Overview

• Show how effective cross-institutional research data management can increase efficiency and improve the quality of research

• Understand and identify requirements for a data-driven research infrastructure in the Structural Sciences (physical science experiments)
  – Examine localised data management practices
  – Investigate data management infrastructure in large centralised facilities

• Scale and complexity: small laboratory to institutional installation to large scale facilities e.g. Diamond Light Source & ISIS, STFC

• Interdisciplinary issues: research across domain boundaries

• Data lifecycle: data flows and data transformations over time
Generalised Issues

- Basic requirement for data storage and backup facilities to sophisticated needs such as structuring and linking together of data
- Management of intermediate, derived and results data a major issue
- Contextual information is not routinely captured
- Processing pipeline is dependent on a suite of software
- The actual workflow or processing pipeline is not routinely recorded
- Need for adequate metadata and contextual information to support:
  - Maintenance and management; Linking together of all data associated with an experiment; Referencing and citation; Authenticity; Integrity; Provenance; Discovery, search and retrieval; Curation and preservation; IPR, embargo and access management; Interoperability and data exchange
- Simplification of inter-organisational communications and tracking, referencing and citation of datasets
  - Unique persistent identifiers
  - Standardised Experiment Risk Assessment forms
- Solutions should be as non-intrusive as possible
An Integrated Service Approach

• Considerable variation in data management requirements across differing scales of science
• Individual researcher, group, department, institution, facilities all working within their own frameworks
• Merit in adopting an integrated framework which caters for all scales of science:
  – Aggregation and/or cross-searching of related datasets
  – Efficient exchange, reuse and repurposing of data across disciplinary boundaries
  – Data mining to identify patterns or trends
• I2S2 Integrated information model aims to:
  – Support the scientific research activity lifecycle model
  – Capture processes and provenance information
  – Streamline flow of metadata, administrative information and experiment data across organisations
  – Interoperate with and complement existing models and frameworks
Benefits - Background

• Based on two use cases:
  – Prof Martin Dove, Earth Sciences, University of Cambridge – researcher perspective
  – Dr Simon Coles, National Crystallography Service, Southampton – service perspective

• Methods developed/enhanced and used:
  – I2S2 Research Activity Lifecycle Model and KRDS Benefits Taxonomy
  – Charles Beagrie Ltd. Value Chain and Impact Analysis Tool
  – Value and Impact elaborated for each perspective by Neil Beagrie working jointly with Martin and Simon
Key Benefits Identified

• Impact and value for researcher (qualitative):
  – research effectiveness – reduced time latency for accessing data sets (24hrs+ down to 5-10mins)
  – disseminating research methods – documented datasets accessible for remote training and learning by (many) new users
  – enhanced research tools – more quality datasets for developers testing and improving tools (software, algorithms, methodologies etc.)

• Impact and value for central service (quantitative):
  – research and facility efficiency – time savings aggregated over many samples /experiments /researchers at facility
  – visibility and security of datasets – increased citation and effectiveness of research in the long-term
  – less likelihood of errors – in data exchange chain between researchers and various facilities (e.g. safety and administrative information)
Benefits Conclusions

• Researcher and Service perceptions of benefits can be and often are different but complementary
• For I2S2, both are positive on benefits that would accrue from implementation
• Impact cannot always be measured within timeframe of project – where appropriate we have established benchmarks against which future progress can be measured
• Finally a similarity to measuring impact in Research Excellence Framework (REF) – a valuable experience for partners given future landscape of research assessment

Diamond Light Source (DLS), Science & Technology Facilities Council, UK
Project Team

- Liz Lyon (UKOLN, University of Bath & Digital Curation Centre)
- Manjula Patel (UKOLN, University of Bath & Digital Curation Centre)
- Simon Coles (EPSRC National Crystallography Centre, University of Southampton)
- Neil Beagrie (Charles Beagrie Ltd.)
- Brian Matthews (Science & Technology Facilities Council)
- Erica Yang (Science & Technology Facilities Council)
- Martin Dove (Earth Sciences, University of Cambridge)
- Peter Murray-Rust (Chemistry, University of Cambridge)

m.patel@ukoln.ac.uk
http://www.ukoln.ac.uk/projects/I2S2/