The Research Data Challenge

Niall O’Loughlin, Research & Enterprise Services, Newcastle University

Aim
- Volume and complexity of research data increasing at significant pace
- Corresponding growth in necessity of managing and storing data properly
- Public funders now requiring data from research they fund made publicly available
- Re-use and result verification possible

How?
- Gather requirements via online survey and semi-structured interviews
- Review policies
- Review existing research information management systems
- Develop research metadata catalogue

Outcomes
Key findings from online survey:
- 61% of research projects had data in physical (non-digital) format
- 27% of researchers have only partial back up of data
- 21% of researchers have RDM plan
- 53% of researchers have no specific RDM tool
- 30% of researchers are aware of DPA
- 13% are aware of FOI
- Few RDM training materials exist at Newcastle University

Early interview analysis suggests:
- University should state minimum RDM standards expected and provide basic RDM guidance
- Integrated institutional RDM approach needed
- Researcher RDM flexibility should be supported if in line with good practice
- Ease of integration of future institutional systems/tools and policies, with existing local/research group tools/systems/policies required
- Better promotion/awareness of RDM-related operating standards, tools, and expertise together with available national services
- Guidance on managing archiving of research data
- Tools/guidance facilitating individual research group data sharing/collaboration (internally within and external to institution)
- Tools for facilitating discovery of research data as research output alongside web profiles/publications

Policy analysis resulted in two outputs:
1. Draft RDM policy consisting of 11 general, high level, static RDM policy principles
2. Draft flexible, dynamic RDM code of good practice: indexing existing University guidance designed to act as toolkit to assist staff in complying with policy principles

Document(s) currently being reviewed by University staff and relevant committees.
- Current University research information management systems cannot handle research data
- A ‘Research metadata catalogue system’ will take feeds of existing project metadata from other systems (Figure 2)
- This will derive appropriate metadata from the project and publications, augmented and exported to an open searchable web-based interface (API) (machine and person readable)

What has worked well?
- Using existing cross-service networks
- Informative and enthusiastic project group members
- Inclusion of two academic units ensures academic credibility and stakeholder engagement
- High level buy-in via University and Faculty Research Committees from an early stage
- Well-informed senior representatives on Steering Group
- High academic engagement due to data storage frustrations
- Regular dissemination via University formal channels

Benefits
- Assessment & trialling of software and process tools
- Ratified policy and code of good practice
- Profile of RDM and supporting services raised
- Direct support for non standard data management in research projects e.g. medical imaging

- Research metadata catalogue will make data both accessible and discoverable enabling re-use
- Embedding of ‘data tagging’ process using existing information and workflows will provide an easy interface for users
- A fully costed business case and report to the University identifying areas where current RDM guidance is insufficient and will recommend improvements
- Assessment of existing & trialling of new training and user support courses
- Compliance with EPSRC’s policy on RDM

Figure 1. Bar chart plots of percentage responses from projects responding to online survey

Figure 2. Schematic diagram of data flows to and from the research metadata catalogue system