A university-wide framework for managing our research data: establishing a Research Data Storage & Management (RDS&RDM) Service

Brief description of the paper
The paper provides an outline business case for PSG to use to decide whether to invest recurrently in a University-wide Research Data Management & Storage Service. The paper has been submitted to PSG as an appendix to the IS Plan for FY2012-13.

Action requested
Library Committee is asked to provide its views on the proposal in terms of need and approach to assist in shaping the work over the coming months.

Resource implications
Yes, IS will underwrite the cost of the first tranche of data storage and the staffing needs of the RDM Service, using accumulated indirects from external grants. Anticipated cost ~£1M one-off, and £250k recurrent.

Risk Assessment
Does the paper include a risk analysis? Not directly, although implicit in the business case.

Equality and Diversity
Does the paper have equality and diversity implications? No

Freedom of Information
Can this paper be included in open business? Yes

Originator of the paper
Jeff Haywood
Vice Principal Knowledge Management and Planning

23 February 2012
A university-wide framework for managing our research data: establishing a Research Data Storage & Management (RDS&RDM) Service

Summary
We propose to develop a university-wide service to provide free-at-point-of-use, adequate storage for research data that will be sufficient for the needs of ~90% of academic staff, researchers and PGRs. It will backed up by a range of services, including training and advice in best practice in research data management. The service will also maintain a University register of data assets, whether held locally or in recognised subject repositories. We will ensure alignment with research grant reporting and compliance. Partnership with Schools will be maintained at all times, and governance for the service will reflect that aim.

IS will fund the initial start-up (18 month) costs of the service, and some of the recurrent costs, from indirects on its own grant-funded activities. Part of the first phase will be to develop a fuller sustainability plan.

What is an RDS & RDM Service?
Digital research data storage and management are required by all academic staff, post-docs and PGRs at the University of Edinburgh as a normal part of their research activities. Sometimes storage is supplied by themselves (physically or in the cloud), sometimes by their research group, School or College, or by others with whom they are collaborating. This may not be, overall, the most effective and efficient approach, and is most likely to be least robust for solo researchers or those with limited technical backup. Training in data management can come from various sources, but is often self-taught.

Data fall into different categories, that we might define as: currently live, recently live and may be re-accessed, long term stored, or archived, and they can be private to an individual, shared within a group, restricted public access or fully open to the world. Funded research produces data that may be required to be made open to the world, sometimes via named repositories, and in some cases for defined periods of time that may well go beyond the presence of an individual researcher within the University. All these activities require hardware and software for data storage/search/preservation/access, but underpinning everything is a need for awareness and training for researchers and support service staff, to enable them to handle their data in the most efficient and secure manner through research data plans from project inception. Aligning an RDS/RDM service with key points in the research lifecycle is essential, for example in the project proposal stage for planning, in the reporting stage for compliance to conditions of grant, which may include a data deposit step.

Why do we need to do something?
For example, while institutions need compliance assurance and protection of their precious assets, researchers need anytime/place/device storage for their data and help with organising and curating them. Funders are raising the stakes: the recent EPSRC announcement of its requirement for data funded by it to be made available for 10 years after the last access, and the expectation of compliance audits, puts pressure on PIs and HEIs alike. Other major funders are known to be reviewing their stance. Local, hand-crafted solutions are not likely to be efficient, nor in many cases effective, especially for the most interesting data which could be accessed long after researchers have left. The University
also needs to know where its data assets are stored, locally and remotely, including, those placed in subject repositories off-campus. When the Arts & Humanities Data Service (AHDS) at KCL closed a few years ago, it returned datasets to the originating universities; almost all of them were unaware that these had been deposited with AHDS. Recent FOI requests for research data or communications about them add a further challenge at individual and institutional levels. Finally, UKRIO good practice guidelines, which the University has adopted, require much of the work proposed here to enable us to be compliant.

It has become clear to many of us working in this field that unless we provide ‘wins’ at each level within the University, we are likely to fail. The sorts of questions that we must address, and to which we must provide satisfactory solutions, are shown in Fig 1 at the end of this paper.

What do we propose?
The complexity of the research data problem, coupled with the dramatically increasing volume of digital data in every subject area, is compelling all universities to seek ways to a) store data securely and reliably and b) manage the process in the most efficient, compliant and cost-effective ways. We propose a university-wide service that offers integrated RDS and RDM. It would place us at the (fore)front of our peers in this key field, in terms of both scale and integration/coordination.

The proposal builds on our strengths. We have made good progress, in that we have a top-level research data management policy, approved by Court (URL), that is being copied and adopted widely as a model of its kind. The policy was deliberately aspirational, and the proposed services will start us on the road to implementation. We have good partnerships between support services and academic units, developed through earlier exploratory work, for example in the Research Data Audit and MANTRA training project. We have the expertise of DCC and EDINA. We have proven central capability to deliver research services, including large-scale data services.

What will the RDS&RDM service offer from summer 2012?

RDS: 0.5Tbyte storage for all academic and research staff, and PGR students (baseline service - no charge – possibility of group sharing); chargeable storage for individuals and groups above the baseline; platform-independent, anywhere access to data; automatic backup; option to move data to archive;

RDM: service elements will include: templates for development of research data plans; good practice guides at subject level for data storage and management; an advisory service (both online and in person) covering compliance and IPR; integration with the PURE service to provide a common outlet for publications and datasets; the establishment of a university data registry; advice on metadata standards for short-term and long-term archival preservation; a full training programme for PGR and ECRs in research data management planning;

Governance: the research data management and storage service will involve stakeholders from across Schools, Colleges and Support Services (eg Records Management, ERI, IAD). A Board will be established to guide the detailed planning, implementation and service
quality aspects of the service, and will contain key stakeholder. It will be chaired by a senior member of academic staff.

Operational responsibility will lie with IS, who will ensure coordination with Colleges, Schools and other Support Services.

**What benefits will accrue?**
Safer storage for all; training and advice, with subject-specific variants; managed access, from closed, through controlled to open; ability to link to UK and international developments (eg UK HE Cloud); a register of all data assets, and audit trails; support for research data publication; the most cost-effective solutions; involvement of researchers as well as support staff, and partnership to enable the service to adapt to changing research behaviours over time; ability to track citation and other impact factors for datasets; increased research impact through better visibility of outputs.

**What will it cost, short term and in steady-state?**
We estimate that the initial start-up costs will be approx. £1M for hardware, software, and School and IS staff. The likely steady-state recurrent cost will be of the order of £250-450k pa, depending on the scale of the services.

The bulk of the initial funding will be provided by IS from accumulated indirects from JISC-funded activities, and the recurrent from these will continue to be invested in this area. DCC and EDINA are able to contribute directly from their reserves and from current UMF-funded activities. Additional sums will be derived from direct and indirect charges to research, especially where large demands are made on the services (eg very large data storage, specific help in managing large or complex datasets).

At the present time, several of our peers are tackling this problem on a full institutional basis, and we are working with them through the JISC Research Data Management Programme to share expertise and experiences. Each university is investing its own funding to stay in the game on top of JISC and other central sources. In Scotland, SFC investment is well below that of HEFCE at the present time, and may continue to be so.

Part of this work is exploratory, but we consider that we have to proceed incrementally, within the funding that we can acquire and prioritising work accordingly, to ensure that we do not fall behind, nationally or internationally.
Where do I keep my fieldwork data safely, as I travel home?

How can I best keep years worth of research data secure and accessible for when I and others need to re-use it?

How do we ensure we have access to our research data after some of the team have left?

How can our research collaborations share data, and make them available once complete?

How do we ensure compliance to funders' requirement for several years of open access to data?

PhD student

research team

university

supra-university

LEVEL

How should I manage all my research data, safely, securely & systematically?

What are Data Management Plans that my funders require?

How do we publish research data to gain maximum credits?

Is there an easy supported data space to promote collaborative working?