Louise Corti
UK Data Archive, University of Essex

University of Essex Sustainability Plan for data management infrastructure

Sustainability of successful systems and infrastructure is dependent upon investment, providing capacity in terms of hardware, software and people. In the case of research data management on a local scale, beneficial activities that enable high quality data management and sharing are best integrated into local routine activities.

During the course of this project the following activities that are seen as having long-term benefits and that are mostly sustainable without any significant additional investment are:

1. Data Management Policy. The University approved a Research Data Policy and drew up a basic route map for data management and sharing that will be taken forward for implementation. The two committees that own this document and route map, the University Strategy Committee and the ICT Steering Group, will review implementation progress on an annual basis and will undertake consultation with local stakeholders at key points. The University is looking to implement a new research administration that will help join up various local research support activities, including costing and monitoring data management and sharing obligations for funded projects

2. Capacity building of REO staff in appreciation of costing data management and sharing and checking data management plans;

3. Capacity building of ISS staff to consider providing a costed model for short to medium-term data storage which research grants can use at the time of grant application;

4. Greater shared understanding of research data management issues across the University research support offices and services with a willingness to create a unified infrastructure;

5. Regular annual awareness and training events aimed at either all staff or trainee/early career staff: sharing your data - ISS and UK Data Archive; Legal and Ethical Issues in Data Sharing – UK Data Archive and REO; planning and costing and data management for grants – REO and UK Data Archive. This is likely to be integrated with research integrity training, including online learning provision, and will look to integrate some of the online learning activities developed by this project;

6. On-going local advice and support from UK Data Archive expert staff, funded under the ESRC's UK Data Service for 5 years from 2012, to support the ESRC Data Policy and engage with anyone in the UK creating social science data that may be of use to researchers and policy makers;

7. Ongoing consultation by the University's IT Steering Group on creating a data storage infrastructure and providing advice for researchers on the ISS website. ISS will work to ensure that the costs of research data storage are coordinated and federated where possible and that explicit contracts for hosting data for research groups are formulated.
   These activities will be carried forward as standard University research support practice.

Eprints app

Under the project an Eprints app, ReCollect, for research data was created and has been added to the Eprints Bazaar where it will be maintained by the University of Southampton. The UK Data Archive has agreed to support queries about the App from other UK institutions for the next 12 months. The project already gained testing and interest from the Universities of Southampton and Leeds.

Priorities in a time of austerity

The University is in a time where funding for new more experimental activities that do not constitute part of core mandatory push factors is extremely limited. For the Essex Research and Enterprise Office, significant investment has gone into two non-routine activities: establishing REF systems and reporting; and setting up a publication repository, leaving little capacity for much else. The University backed this 18 month JISC bid by contributing overheads and support staff time and expertise, by engaging with its research community, and by initiating a Data Management Policy and a route map for implementation.

However, the University is uncertain as to whether a single-track investment in local data infrastructure is the best way forward and that a federated approach with, for example, their 94 group colleagues might be more efficient and less cost-intensive. It needs to be recognised that there is already both formal mandatory data sharing policies in place that are working and national expertise in data centres who already deal with sharing research data. Smaller Universities, such as Essex, would be better using national shared infrastructure services, like UMF brokerage Services for cloud storage and possibly specialist off-site data appraisal and advice on preparation, such as that provided by the UK Data Archive for all ESRC research data. Any longer-term preservation comes at a huge price and should utilise a shared service. The University is aware that some institutions are looking at commercial enterprises to provide 'dark archiving' services, seemingly at great cost.

Capacity building and compliance

Continuing activities in this area are important for the following reasons:

1. Improving local data management planning knowledge to meet funders' application requirements;
2. Improving local data sharing knowledge to meet funders' award requirements;
3. Ensuring local compliance with the Freedom of Information Act and Environmental Information Regulations;
4. Improving early career skills in data management and sharing.

This capacity will be delivered in the future via core local centralised workshops and online modules; local guidance on website, links to other information providers and external training resources.

Repository options

While a small scale-data environment, using the Eprints Open Source software has been proven in this project, the University is uncertain as to how it might support this system in the near future. The current Essex Research Data system will be left up as a pilot system for the next 6 months and can be picked up if the need arises. There are two options for possibly implementation going forward:

1. The ReCollect Eprints system can act as a metadata storage system once the EPSRC Policy becomes mandatory. A metadata system would require minimal investment as it is set up to go. It is suggested that this would require 5% of junior staff REO staff time (adapting fields, checking metadata, publishing the record and linking to research outputs but excluding chasing up PIs), and 10% of a junior staff member in ISS to maintain the server and software updates. It would need to be integrated with the new research administration system being planned for post 2014.

2. A full-fledged local data storage repository system would require more significantly more investment in staff time and hardware:
    a) Repository manager: adapting and checking metadata, data appraisal, checking data for legal and ethical issues, publishing records; linking to research outputs, negotiation and contact with PIs, checking contractual obligations; advocacy activities and integration with other research integrity activities: full-time junior staff REO staff time;
    b) Server and system maintenance: software updates, configuration, importing regular data sources, link to non-Eprints storage of larger-scale data: 20% of a junior staff member in ISS;
    c) Longer-term local storage capacity, the cost of which cannot be estimated at this time as total data capacity is unknown;
    d) Preservation activities need to be considered and this would come at an extra cost, making use of contracted shared services.

The University has put both these on hold until it becomes mandated by RCUK or publishers to provide such services. As far as its own researchers are concerned, adequate facilities already exist to enable sharing of data where it is contractually required and a significant local need is not yet evident. Any university-level approach would need to be supported by shared services across groups of universities, to demonstrate economies of scale.