# JISC

# Research Data @ Essex Final Report

| Project Information | | |
|---|---|---|
| **Project Identifier** | *To be completed by JISC* | |
| **Project Title** | Research Data @ Essex | |
| **Project Hashtag** | rdessex | |
| **Start Date** | 10/10/2011 | **End Date** 31/3/2013 |
| **Lead Institution** | UK Data Archive, University of Essex | |
| **Project Director** | Louise Corti | |
| **Project Manager** | Louise Corti/Veerle Van den Eynden | |
| **Contact email** | tensom@essex.ac.uk; corti@essex.ac.uk ; veerle@essex.ac.uk | |
| **Partner Institutions** | University of Essex | |
| **Project Web URL** | http://www.data-archive.ac.uk/create-manage/projects/rd-essex?index=0 | |
| **Programme Name** | *Managing Research Data 2011-2013* | |
| **Programme Manager** | Simon Hodson | |

| Document Information | | |
|---|---|---|
| **Author(s)** | Veerle Van den Eynden, Tom Ensom, Louise Corti | |
| **Project Role(s)** | Project manager, Project officer, Project director | |
| **Date** | 08/05/2013 | **Filename** ResearchDataEssex_finalreport_01 |
| **URL** | *http://www.data-archive.ac.uk/create-manage/projects/rd-essex?index=1* | |
| **Access** | This report is for general dissemination | |

| Document History | | |
|---|---|---|
| **Version** | **Date** | **Comments** |
| 01 | 08/5/2013 | Final report submitted to JISC |
| | | |
| | | |

# Table of Contents

## 1. Acknowledgements

This project was funded under the JISC Managing Research Data programme 2011-2013. The project team, led by the UK Data Archive, included staff from Information Systems & Services (ISS), the Research and Enterprise Office (REO) and Records Management at the University of Essex. We worked closely with researchers in four of the university's research-rich departments: Biological Sciences, Essex Business School, Language & Linguistics and Computing & Electronic Systems. The EPrints team at the University of Southampton provided valuable input into the development of the data repository, as well as the development of the ReCollect app. A Steering Committee composed of representatives from each of the participating entities (including the researcher community) oversaw the strategic processes of the project.

## 2. Project Summary

The Research Data @Essex project, funded under the JISC Managing Research Data Programme, piloted the development of a research data management and sharing infrastructure for the University of Essex. We developed an institutional research data repository, Essex Research Data (built on the EPrints institutional repository software; and in collaboration with the EPrints team at Southampton released the data repository App 'ReCollect' via the EPrints bazaar. This App allows institutions to set up their own EPrints-based data repository.

Extensive in-house customisation was done to make the EPrints software - initially developed for hosting research publications - more suitable for storing and presenting research data. Development has focused on expanding the EPrints metadata profile to allow the capture of detail necessary for describing a wide diversity of research data collections, while also meeting standards relevant to UK Higher Education. The metadata profile adopted is based on DataCite (2012) and the European INSPIRE metadata schema (INSPIRE, 2009), and is compliant with the international specification Data Documentation Initiative 2.1. (DDI, 2011), used by the UK Data Archive and every social science data archive around the world. In addition, the user interface was adapted to showcase complex collections consisting of many files and associated metadata, alongside a refined deposit workflow. Work also included the evaluation and selection of suitable licensing agreements for research data; methodologies for attaching persistent identifiers to the data collections, including DataCite Digital Object Identifiers; and the development of a user guide for preparing and depositing research data in such institutional repository.

The project team, led by the UK Data Archive, included staff from Information Systems & Services (ISS), the Research and Enterprise Office (REO) and Records Management at the University of Essex. We worked closely with four of the university's research-rich departments: Biological Sciences, Essex Business School, Language & Linguistics and Computing & Electronic Systems. This involved evaluating current research data management and sharing practices, as well as ingesting representative sample data into the pilot repository, including proteomics and bio-imaging data, socio-linguistic data, artificial intelligence data and management research data. Close working with researchers ensured ample testing and polishing of the software.

Besides the repository development, the project also developed an institutional data management policy for the university and a roadmap for compliance with the UK's Engineering and Physical Sciences Research Council (EPSRC) policy framework on research data. Data management guidance and training are also provided for the university's research community.

## 3. Main Body of Report

## *3.1 Project Outputs and Outcomes*

| Output / Outcome Type *(e.g. report, publication, software, knowledge built)* | Brief Description and URLs (where applicable) **All outputs are available on the Research Data @ Essex project outputs webpage** |
|---|---|
| Data repository | Pilot research data repository, Essex Research Data, for the University of Essex, built on EPrints repository software. |
| Repository user guide | A user guide for preparing and depositing research data in the institutional data repository |
| Repository policies | Relevant data repository policies |
| Software application | Data repository app 'ReCollect' available via the EPrints bazaar. This app allows institutions to set up their own EPrints-based data repository with expanded metadata profile for describing research data and a redesigned data catalogue for presenting complex collections. |
| Report | Repository development and ingest report, describing development rationale, ingest testing with sample data and the consultation with Essex researchers and external specialists. |
| Metadata profile and mapping | Expanded metadata profile based on standard metadata schemas such as DataCite, INSPIRE and the Archive's DDI2.1; a metadata mapping table shows how metadata map across those standards. |
| Policy | The University of Essex Data Management Policy and route map for compliance with the EPSRC Policy Framework were developed as part of this project and published in March 2013. |
| Guidance | Data management planning guidance for University of Essex researchers |
| Training | A regular programme of Research Data Management workshops takes place for University of Essex researchers as part of the University of Essex Learning and Development programme. Also general UK Data Archive organized research data management workshops can be attended by University of Essex (UoE) staff and students. Training events held: <ul><li>Going digital - Looking after and managing your digital research data on 24 April 2013 (6 UoE PhD students attending), organised as part of the AHRC Going Digital programme</li><li>Sharing your Research Data – Ethical & Legal Issues on 24 January 2013 (7 UoE staff attended)</li><li>Looking after your Research Data on 29 February 2012 (21 UoE postgraduate students attended), 5 December 2012 (9 UoE postgraduate students attended) and 19 February 2013 (5 UoE staff attended)</li><li>Managing your Ph.D. research data on 22 November 2012 (6 UoE postgraduate students attended)</li><li>Looking after and managing your research data: an ESRC Advanced Training Network course on 26-27 April 2012 (1 UoE postgraduate student attended)</li></ul> |
| Report | Report, summary results and questionnaire form of the research data management survey held at the University of Essex during February 2013 to examine current practices and |

| | needs with regards to data management planning, data sharing and a data repository. |
|---|---|
| Report | **Research Data Management at the University of Essex** describes the findings of the assessment of research data management practices within four departments of the University of Essex, including a Data Inventory Form, derived from an earlier data management practices assessment carried out with research centres as part of the Archive's DMP-ESRC project in 2011, and also based on the Data Asset Framework methodology. |
| Report | University of Essex sustainability plan for data management infrastructure |
| Blog | Project blog with information about project activities and relevant topical discussions |

## 3.2 How did you go about achieving your outputs / outcomes?

The Research Data @ Essex project worked to develop a research data management and sharing infrastructure for the University of Essex, to address the originally set out objectives to scope and develop a pilot research data management infrastructure for the University of Essex to strengthen research capacity at Essex, widen the availability of data resources, and improve local attitudes towards managing and sharing research assets.

The following outcomes were achieved:

- policy & advocacy:
  - o development of a Research Data Management (RDM) policy and ESPRC policy framework roadmap for the University of Essex
  - o sustainability report for research data management infrastructure
- centralised data management support & training for the university's researchers
  - o data management planning guidance and support
  - o ongoing programme of data management training
- development of a pilot data repository from EPrints software
  - o user interface and workflow customisations of EPrints
  - o metadata profile & standards (DDI, INSPIRE)
  - o repository policies
  - o digital object identifiers for data collections


The various university partners, the UK Data Archive (UKDA), Information Systems & Services (ISS), the Research and Enterprise Office (REO) and Records Management worked together in the project as follows. REO and ISS developed the data policy, in communication with other partners. UKDA with REO developed the support and training framework. UKDA and ISS jointly developed the pilot data repository. UKDA with the project steering group developed the sustainability plan.

Activities and achievements were solidly underpinned by an assessment of current research data management and sharing practices as well as future data management needs and expectations, through in-depth work with four pilot departments and its researchers and research directors, and through a generic staff survey. The participating research-rich departments were: Biological Sciences, Essex Business School, Language & Linguistics and Computing & Electronic Systems. The work with the four departments also ensured that all developments related to the data repository were in line with researchers' research practices and needs. Participating researchers provided diverse sample research data collections for ingesting into the pilot repository: proteomics and bio-imaging data,

socio-linguistic data, artificial intelligence data and management research data. Through testing they help ensure that the developed repository was fit for purpose and user-friendly.

## 3.2.1 Data practices evaluation with pilot department

Data management interviews to assess current research data management and sharing practices were held with the four research leaders of participating departments, as well as with 13 researchers actively collecting and managing research data.

Research director interviews focused on research data policy, infrastructure and higher level perspectives of the incentives and sanctions in place to push forward the integration of data management. The role of research directors includes dealing with local data infrastructure needs such as facilities and technical support, coordinating research funding and research proposals ethical review for staff and student research projects, research appointments and annual staff reviews, coordinating Research Excellence Framework returns and monitoring on-going research activity.

Researcher interviews used a Data Asset Framework based methodology (DCC, 2012), that was adapted from similar data management assessments carried out by the same team within research centres (rather than an entire university) as part of the JISC-funded DMP-ESRC project (Horton et al., 2011). Creating a data assets inventory is a widely used method of assessing institutional research data assets. The use of semi-structured interviews gave researchers the opportunity to discuss their own specific concerns and challenges, or suggest what they would like to see the university provide.

The topics covered during interviews, which were chosen based on previous projects within the research data management community, were:

- data management strategy
- data documentation
- file formats
- ethics and consent in research
- copyright and IPR
- data storage and backup
- file naming and versioning
- sharing and reuse of data
- data security and destruction

On 13 December 2011 an informal data managers meeting was held with invited research directors, to discuss the option of establishing a regular data managers forum for the university. Whilst considering research data management to be important in research, the group felt that a formal forum with regular meetings would not be useful at this early stage of data management infrastructure development. Instead, research directors were keen to take on a local information role and to meet when critical matters would arise. The Research and Enterprise Office will take forward future meetings with research directors to discuss practical implications of the EPSRC policy framework and the university's data policy implementation.

## 3.2.2 Research data management staff survey

An online staff survey was held amongst all academic staff of the University of Essex during February 2013 to inform development of a sustainability plan and to direct future training activities.

The aim of this survey was to:

- raise and assess awareness amongst research staff about data management planning requirements and practices;
- assess current practices and support needs with regards data management and sharing

planning;

- assess current data sharing and publishing levels by researchers;

- assess whether researchers have a need for a data repository at the university;

- elicit feedback and testing of the developed pilot data repository "Essex Research Data", to ensure that the repository is fit for purpose for the university's research community.

The survey was developed based on research data management expertise at the UK Data Archive, and also drew inspiration from a recent similar research data management survey held at Oxford University by the DaMaRO project.

Fifty-five responses from researchers at the University of Essex were received, from across 16 of the university's 23 departments. Most responses were received from researchers in biological sciences (20%), language and linguistics (15%) and business sciences (15%). The 55 respondents represent 13 percent of all academic research staff at the university.

A report, survey questionnaire and summary report of all responses have been published.

## 3.2.3 Policy and advocacy

The Research and Enterprise Office (REO), together with ISS and RM took forward the development of a University of Essex Research Data Management Policy. The policy, which consists of 11 core principles, sets out how research data is managed throughout the data lifecycle to ensure transparency and appropriate access, supporting the UK Research Council's requirements and Freedom of Information legislation. A route map identifies and maps the steps needed to be taken in order to achieve full compliance with the requirements of the EPSRC. This route map considers: data management support, data storage, data register, progress reporting, implementation resources and compliance review.

The policy and route map for policy implementation were approved by the university's Research Strategy Committee (RSC) on 26 June 2012 and by the ICT Steering Group (ISG) on 14 November 2012. The policy and route map were published by the university in March 2013.

REO will lead the policy implementation, which will be overseen by the RSC and ISG. Annual implementation reports will be presented to RSC and ISG starting in March 2014.

A sustainability plan for research data management infrastructure for the University of Essex was developed to inform REO, RSC and ISG.

## 3.2.4 Support and training for researchers

The UK Data Archive now runs regular data management workshop for University of Essex research staff and students, as part of the university's Learning and Development programme for staff. ISS have been and will continue to be involved in running these workshops, providing advice on local data storage, back up, collaborative data sharing and security issues. Also general UK Data Archive organized research data management workshops can be attended by University of Essex (UoE) staff and students.

Whilst bespoke training was offered to the collaborating departments, none have taken up the offer but instead point staff and students to the courses offered via the Learning and Development programme.

In the course of this project, the following training events were held:

- Going digital - Looking after and managing your digital research data on 24 April 2013 (6 Essex PhD students attending), organised as part of the AHRC Going Digital programme
- Sharing your Research Data – Ethical & Legal Issues on 24 January 2013 (7 staff attended)
- Looking after your Research Data on 29 February 2012 (21 postgraduate students attended), 5 December 2012 (9 postgraduate students attended) and 19 February 2013 (5 staff attended)
- Managing your Ph.D. research data in November 2012 (6 postgraduate students attending)

- Looking after and managing your research data: an ESRC Advanced Training Network course on 26-27 April 2012 (1 Essex postgraduate student attending)

Research funder data management planning requirements and how best to support researchers was discussed with REO staff that support research grant submissions and award management. Most research grant applications being submitted via the REO are for funding from the Leverhulme Trust, the British Academy and UK Research Councils. Up until now REO staff have not been faced with problems for the development of data management plans. A few requests for guidance were passed on to UK Data Archive staff.

Data management planning guidance has been published on the REO website. This guidance provides a brief overview of relevant topics, then points to the well-established UK Data Archive Create & Manage Data website for more specific guidance and also provides contact details when help is needed.

REO is interested in the Archive further developing its data management costing tool that was developed by the JISC-funded DMP-ESRC project, and to incorporate it on the REO's grants information web pages. Grant applications currently tend not to include data management costs. The Archive will develop the tool further in preparation for a planned grant writing workshop (date not yet confirmed), to be held jointly with REO.

## 3.2.5  Pilot data repository development

A pilot data repository 'Essex Research Data' was developed via a collaboration between ISS and the UK Data Archive, to provide a means for the university's researchers to facilitate the sharing of research data, both within the university and with the wider community. In further collaboration with the EPrints and DataPool project teams at the University of Southampton, the data repository elements were converted into a data app (or plugin) for EPrints called ReCollect, which is made available via the EPrints bazaar.

This data repository was developed from EPrints software, which is primarily used to build publication repositories. The University of Essex research base is broad, and many data types are produced – from large scale proteomics to business management spreadsheets. It was an essential requirement that any system we developed be generic enough to store, describe and present data from any of the disciplines represented at Essex. We were also keen to ensure that the deposit process did not become overly complex for a user, in order to minimise uptake barriers. Taken together, these two factors involved a necessary compromise between ease of deposit and the need for sufficient information to fully enable re-use.

We adapted EPrints by extending the metadata profile and the user interface for the deposit and display of data collections, adapted licensing agreements suitable for research data and considered how digital object identifiers would best be assigned. All development was carried out in participation with a small group of researchers from different departments at the University, who provided test data for ingest, gave functionality and needs suggestions from their unique perspectives and gave feedback during demonstrations of the ingest and dissemination process. This was followed by the test ingest of diverse kinds of real-case data collections and evaluation by 18 internal and external evaluators. The latter were from the JISC Research data management community, including developers, repository administrators, data librarians and other information professionals.

## 3.2.6  Metadata profile for data repository

The basis of the extension of the default EPrints metadata was a three-layer metadata model, based on the UK Data Archive's approach to social science data archiving and also adapted from work by the JISC IDMB project (University of Southampton, 2011) of:

- core metadata (for citation and discovery): author, discipline, project,…

- detail metadata (descriptive and contextual) : collection, methodology,…

- discipline-specific metadata (as additional metadata files):

The core and detail metadata were constructed using elements from the following existing schema, and mapped to DataShare, Edinburgh's digital repository (University of Edinburgh, 2013):

| Schema | Area | Compliance | Reason for use |
|---|---|---|---|
| DataCite (2012) | Data citation | Yes | Minimal mandatory elements; essential if DataCite DOIs are to be minted. Emerging as a *de facto* standard among data repositories |
| INSPIRE (2009) | Geospatial data | Yes | EU standard; generic description; intended for different types of data with geospatial content |
| DDI 2.1 (2011) | Social sciences data | No | Descriptive/contextual metadata beyond scope of above schema e.g. collection methodology, ethics/consent statement |

For subject classification, we used the HESA JACS3 scheme (HESA, 2011) instead of EPrints' Library of Congress classification default, resulting in a much-improved mapping to the disciplines represented at Essex. Also the RCUK scheme of subject categories is another strong candidate.

The final metadata profile for the Essex Research Data Repository was published.

## 3.2.7  User interface and workflow for data repository

The default EPrints user interface (UI) was not found to be able to present data collections containing large numbers of data files and their metadata with sufficient clarity in the data catalogue. The design for a data record also needed to be adapted to be able to show the extended metadata profile for a data collection. We therefore created a custom user interface that groups data and documentation files in a box on the right hand side of the screen. This grouping is made according to the type of file that the user selects during upload, selecting from: data; documentation; additional metadata; readme; or full archive. The latter represents all files within a data collection. Also, a detailed metadata section was added to the bottom of the screen of a data record. All metadata is thereby visible on a single screen (page), while maintaining an uncluttered interface.

Some workflow alterations were implemented to allow for the metadata and user interface extensions. We removed the option of selecting a resource type (article, dataset etc.), as this was redundant. There have also been tweaks to the order in which the fields appear during the deposit process to keep logically related fields together. New and updated help text and guidance has been added, targeted in particular at research data deposits.

## 3.2.8  Other data repository adaptations

In addition to the above, we have created user documentation to assist users with the deposit process and particularly to elucidate optional or conditional elements of the workflow.

We also drafted repository policy documents specifying end user licence conditions and terms and conditions of deposit.

The repository enables uploading of individual data files or zip files that bundle up numerous similar data files, allocation of descriptions, searching in file names and content, and assigning a license to a data collection.

## 3.2.9 Digital Object Identifiers (DOIs) for the repository

We investigated how we could assign permanent identifiers for the EPrints system. The naming convention EPrints uses seems quite suitable and intuitive as a unique identifier in itself, thus GUIDs were felt to be sufficient as long as the Repository system did not change its root URl. A data collection identifier such as "/Essex-EP-105-5" could resolve to a data object.

Following wider discussions at JISC events, further plans to test a DOI system using the DataCite APIs were not pursued as there has been debate and uncertainty as to how useful Datacite DOIs might be in an institutional repository context, as is described on the project blog. During the project we discussed the use of Datacite DOIs with a number of other HEIs in the Programme and all were also 'waiting' until there was a consensus as to the real value of buying into Datacite DOI (and the contractual cost factor permitting).

## 3.2.10 Sample data preparation and ingest

Sample datasets were provided by participating researchers, covering a broad range of sub-disciplines. Each dataset was ingested, including the preparation and upload of files, and the addition of metadata at the data file and data collection level. In addition, we also trialled uploading a selection of datasets held in the ESRC Data Store.

The following test data for ingest into the Essex Research Data Repository have been obtained and ingested:

- Proteomics, mass spectrometry data from tumour tissue samples
- Bio-imaging, high resolution image data collected to examine cellular structure
- Management, Football manager performance exploring managerial succession in an environment of instant public access to performance metrics
- Linguistics, Second language acquisition, audio and transcripts of classroom second language learners
- Sociolinguistics, audio and transcripts of interviews with multiple generations of Indian English speakers

## 3.3 What did you learn?

### 3.3.1 Data practices assessment and survey

The assessment interviews with researchers and research directors in pilot department found that:

- local control and management of data is varying by department and research group

- little awareness exists of formal roles and responsibilities for data governance and storage within departments or the university as a whole

- no centralised support exists currently for longer-term data storage and to provide data access

- little awareness exists of structures and procedures for responding to Freedom of Information requests for research data

- little awareness exists of the forthcoming ESPRC data framework

- no real thought is given to costing longer-term data storage and access into grants

- no staff or student training exists on data management, data storage and data security issues

- data management planning is a new concept to many researchers

- there is a lack of centralised guidance on research data management issues

- an institutional research data repository would be valued by researchers

- data management practices are largely focused on own research needs, without taking potential future data usage into consideration.

Almost all the researchers interviewed expressed concern over how their data are stored and felt it was difficult to exchange or share research in collaborative research, both within the university and with external partners.

The assessment findings have been published in a report. This report provides an interesting snapshot of a university just prior to the emergence of a research data management infrastructure, a snapshot which we suspect bears similarities to other universities in the UK.

During meetings with research leaders, UK Data Archive data management tools and resources such as the data management costing tool, institutional resources library and guidance were showcased.

The research data staff survey found that researchers at the University of Essex generate a mixture of data types, with numerical and textual data being the most prominent ones, besides images and audio data, multimedia data, bibliographic data and geospatial data. Most researchers produce relatively small data files of less than 2 gigabyte per file. During research, active data files are mostly stored on PC or laptop hard disk, in online or cloud storage and on external hard disk. For longer term storage after projects end, researchers rely on hard disks, online and cloud storage and repositories. The main challenges they experience with regards data storage are the sharing and storage of data in collaborative research, version control in collaborative research and storage capacity.

Only a third of those obtaining external grant funding for research have currently been required to complete a data management and sharing plan as part of grant applications. Many are, however, aware that major research funders in their subject domain tend to require data management plans. Fifty percent of respondents do not know where to seek help for completing a data management plan. Those that do know would approach a data centre service or the university's research office. Data management training needs were identified during the survey.

Overall, respondents are fairly familiar with and/or positive about sharing their research data, with researchers having a funder or publisher requirement to share their research data, or a legal requirements to do so. Most considering their data to be an important resource for future research and/or learning.

Forty five percent of researchers have placed data in a data repository, mainly because they are in favor of data sharing, because it is community practice in their research field, or due to a requirement from publishers or funders. The repositories or data centres used are national research council repositories and international disciplinary repositories. Those researchers not used to data sharing are either not aware of any suitable data repository or centre, consider their data not suitable for sharing or do not find the time to prepare data for sharing. Most respondents would consider sharing some of their data in future, either without restrictions, upon request or with known collaborators.

A quarter of respondents would like to deposit research data in a data repository at the University of Essex in future, but only 20 % are (very) likely to need to place data in a repository during 2013, either due to funder or journal requirement, for research excellence purposes, as best long term storage strategy or for the benefit of future research and collaborations.

### 3.3.2  Pilot data repository development

Upload of diverse types of test data provided by collaborating researchers and evaluation and feedback from those researchers as well as from internal and external testers provided important input into the development of the functionality, features, the ingest and metadata submission flow and the look of the pilot repository and therefore also of the ReCollect app. For example, data collections consisting of numerous large volume data files and data collections requiring access regulation due to ethical requirements informed the development strongly.

## *3.4  Immediate Impact*

At the University of Essex the project resulted in the development of an institutional data management policy for the university and a roadmap for compliance with the UK's Engineering and Physical Sciences Research Council (EPSRC) policy framework on research data. In addition, data

management planning guidance has been included on the REO website and data management training workshop are now a regular feature in the university's Learning and Development training programme for staff and students. This provides the basis of a slow increase in data management awareness and skills for researchers at the university. Regular discussions in the course of the project with REO, ISS and research directors also resulted in them being more aware of funders' requirements for data management planning.

The developed pilot institutional research data repository, Essex Research Data, is considered to be a potential future model for a functional data repository for the university, should that become a strategic need. Alternatively, it could be set up as a metadata storage system for EPSCR data policy framework compliance.

At the same time, the data repository provides an excellent model for other institutions and the wider academic community, in particular through the development of the EPrints data app 'ReCollect', which is freely available. This means other institutions can easily install a similar repository for their purposes. The project already gained testing and interest from the Universities of Southampton, Leeds and Bath, as well as the Institute of Science and Technology, Austria. Furthermore, the metadata profile as a separate output has wide re-use potential as a model for other institutions not using EPrints.

Also at the UK Data Archive, the development of a data repository based on EPrints has been seen as promising and funding has been obtained from the ESRC to develop a new ESRC Data Store, used for the self-deposit, management, storage and dissemination of research data from ESRC grants, based on this model.

The following evidence is being gathered to measure immediate and future impact:

| Benefits | Evidence |
|---|---|
| UoE researchers- Increased DM awareness and skills | Nr staff at UoE RDM training and their evaluation / feedback<br>Use of REO DMP guidance and UKDA Create & manage online guidance (Google analytics)<br>Qualitative feedback (staff survey)<br>Requests for advice to REO, ISS, UKDA |
| UoE support staff (REO) – increased DM and DMP awareness, capacity building | Qualitative feedback<br>Number of joint workshops REO-ISS-UKDA<br>New streamlined RDM content on REO website<br>New DM costing content for REO website<br>RDM policy and roadmap approved and used as working document |
| Visibility of data assets | Nr test datasets published<br>Nr people requesting account and uploading test data<br>Interest shown by UoE researchers to publish / ingest their data (staff survey)<br>Nr people accessing pilot repository for testing / use |
| Pilot data repository developed for UoE | Qualitative feedback from researchers of pilot departments on their deposit<br>Consideration, interest and adoption of metadata model by other repositories and EPrints community<br>Adoption of repository system by data centres (UKDA, APA, Austrian Inst Sci)<br>Development of EPrints Data App |

## *3.5 Future Impact*

Various achievements of the project will continue to have an impact for researchers and research data at the university:
- greater shared understanding of research data management issues and a willingness to created a unified infrastructure across the university's research support offices and services;
- data management planning guidance available to researchers;
- ongoing data management training courses;
- data management (planning) advice that can be sought from REO and the UK Data Archive.

Regular annual awareness and training events aimed at staff, early career staff and postgraduate students will continue to be held as part of the university's Learning & Development programme on the topics of:
- Sharing your research data - ISS and UK Data Archive
- Legal and Ethical Issues in Data Sharing – UK Data Archive and REO
- Planning and costing data management for grants – REO and UK Data Archive

This is likely in the future 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.

Ongoing local advice and support will be provided by the UK Data Archive staff, funded under the ESRC's UK Data Service for 5 years from 2012, to support the ESRC Data Policy.

Also support staff will continue to benefit from their increased awareness of critical data management aspects. At an institutional level, the data policy will be implemented by REO. The University is also looking to implement a new research administration system that will help join up various local research support activities, including costing and monitoring data management and sharing obligations for funded project.

Beyond our own institution the availability of the ReCollect data app in the EPrints Bazaar, where it is maintained by the University of Southampton, means that other institutions will be able to easily adopt a data repository system. The UK Data Archive will support queries about the app from other UK institutions for the next 12 months. The pilot Research data Essex repository will remain online as a pilot repository for at least six months. The metadata profile that the app supports has been published separately and would also enable other non-EPrints repositories to implement it.

We will continue to track, via Google Analytics, the use of data management guidance by Essex researchers. Workshop attendance by Essex staff and students, both at workshops within the Learning & Development programme as other data management training workshops by the UK Data Archive is being monitored.

The implementation of the university's data policy will be monitored via annual reports to the Research Strategy Committee and the ICT Steering Group.

The adoption of our developed repository system by other institutions or organisations will be monitored via downloads of the ReCollect app. Other simple plugins from EPrints enable downloads to be measured and also publically viewed if required.

## 4. Conclusions

Overall we achieved what we had expected to achieve. Research data management may not yet be seen as an immediate nor short-term strategic priority for the University of Essex, but various partners from within the university actively engaged in this project meaning that we achieved results on all fronts, albeit sometimes slower than hoped for.

Whilst the case has not yet been made to turn the developed pilot data repository into a functioning repository for the university's research community (who have indicated their interest), the knowhow gained remains present within the university as the staff members who carried out the repository development will continue to be employed within the UK Data Archive and ISS. This means that at a

later stage a data repository can be set up for the university. In the meantime the UK Data Archive will continue to develop the repository into a new ESRC Data Store system.

With regards the wider community, we had an excellent collaboration with both the DataPool project and the EPrints team at the University of Southampton, which resulted in us achieving more than had been planned, namely the publishing of a data app that can be adopted by other institutions. We also gained potential interest from other non-HEI organisations (e.g. APA, the Hungarian Data Archive).

We are confident that the ReCollect metadata profile we developed via intensive mappings, which was successfully tested across disciplines, represents a gold standard for describing research data collections. We hope that any future collection–level research data harvesting systems will consider this as a baseline.

We have developed a comprehensive solution for institutional data storage, built on the back of existing and widely used open source software. Our approach is one that we hope will be adopted (through the ReCollect app or otherwise) by institutions new to the research data management challenge and looking for a low-barrier solution for setup and standards compliance. It can be particularly useful in getting institutional buy in when EPrints is already deployed as a publications repository.

## 5.  Recommendations

With regards the developed repository, there are several gaps in the system that emerged during our work, and we hope the community will agree that tackling these with urgency will be of benefit. For practical reasons, individually tagging many files with their respective metadata is not at present something depositors could be expected to undertake. We would like to see tools/extensions developed to enable a better way of tagging large numbers of files with metadata, or allowing for file level inheritance of metadata from a zip file. In the meantime, we will not be recommending that depositors upload large numbers of files and individual tag them, but rather upload and tag zip files containing them.

Affecting the most repository software are issues with the upload of files over 2GB in size. This seems to be to do with the limitations of browser protocols. One option may be to encourage such files to be hosted elsewhere, and linked to through the EPrints metadata record. There is also the need to extended upload options – for example the use of BitTorrent like protocols.

Towards the University of Essex our recommendation remains to ensure the effective implementation of the data policy and the continuation of the provision of research data management training and guidance, as outlined in section 3.5.

Towards JISC our recommendation would be to support the development of federated cross-institutional data repositories (data infrastructure), as this may be a better solution for many institutions, rather than setting up individual institutional repositories, due to the staff with data expertise that is needed to successfully administer and manage a data repository for long-term sustainability, and the substantial cost of professional curation and long-term preservation.

We also recommend that any future collection–level research data harvesting systems developed in the HEI community consider the ReCollect metadata profile as a baseline. Since it was successfully tested across numerous research disciplines and resulted from mapping across various metadata schemas, it represents a gold standard for describing research data collections.

## 6.  Implications for the future

We encourage other institutions to adopt and build upon the developed EPrints data repository system, metadata profile, licencing model and fucntionality.

The ReCollect metadata profile and metadata mapping can strongly inform the metadata activities of JISC's planned Research Data Registry project.

The ReCollect app is available via the EPrints bazaar.

The findings and outputs of the projects will continue to be available from the project outputs webpage.

## 7. References

DataCite (2012) DataCite Metadata Schema 2.2 [http://schema.datacite.org/]

Data Documentation Initiative (2011). DDI Codebook 2.1 specification [http://www.ddialliance.org/Specification/DDI-Codebook/2.1]

DCC (2012) Data Asset Framework Implementation Guide. [http://www.data-audit.eu/]

EPSRC (2011) EPSRC Policy Framework on Research Data. [http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx]

HESA (2011). Higher Education Statistics Agency, JACS3 Listing v 1.2 [http://www.hesa.ac.uk/content/view/1805/296/]

Horton, L., Van den Eynden, V., Corti, L. and Bishop, L. (2011) Data Management Recommendations for Research Centres and Programmes. UK Data Archive, Colchester. [http://data-archive.ac.uk/media/257765/ukda_datamanagementrecommendations_centresprogrammes.pdf]

INSPIRE (2009). INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. v1.1. [http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20090218.pdf]

University of Edinburgh (2013). DataShare [http://datashare.is.ed.ac.uk/]

University of Essex (2013). Recollect. EPrints Bazaar [http://bazaar.eprints.org/280/]

University of Southampton 2011. Institutional Data management Blueprint project [http://www.southamptondata.org/]