RESEARCH DATA@ESSEX

RESEARCH DATA MANAGEMENT TEAM
UK DATA ARCHIVE, REO and ISS
UNIVERSITY OF ESSEX

DCC ROADSHOW
LONDON, 21 MAY 2012

UK • DATA ARCHIVE

PROJECT AIMS

- JISC funded under the Managing Research Data Programme call, 2011
- We are piloting a research data management infrastructure for the University of Essex
- Builds on the UK Data Archive's existing user support and capacity building work with JISC and ESRC on managing and sharing data
- Building on the University’s recent implementation of Eprints for its outputs repository

PARTNERS

- UK Data Archive
- Research and Enterprise Office (REO)
- Records Management Office
- Information Systems Services (ISS)

INSTITUTIONAL DATA POLICY

- A number of University-wide Data Management Policies have come into force recently
- Essex is slow but is building on the EPSRC Framework document
- Following Edinburgh 10 point policy. Being prepared now by:
  - Research and Enterprise Office (REO)
  - Records Management
  - Information Systems Services (ISS)
- This project is helping push this along: guidance will include scenarios, 'what if I'? Q&A
- Existing Statement on Safeguarding Good Scientific Research Practice
- Records management policies procedures but not so detailed for research data

Existing Essex statement

Documenting results and storing primary data

Researchers are required to keep clear and accurate records of the scientific procedures followed and of the results obtained. This is necessary for a number of reasons, including:

- The records will provide a means of demonstrating proper scientific practice

- The records will be available in case questions are subsequently asked about either the conduct of the research or the results obtained

- The records can ensure that IPR can be protected

- Academic staff are responsible for storing their records securely in an appropriate durable form. The appropriate period for retaining data will depend on circumstances; for some disciplines, the importance and relevance of data can be superseded very rapidly. However, as an example, BBSRC expects data to be securely held for a period of ten years after the completion of a research project. An appropriate period should be set within each department.

THE DATA ARCHIVE'S WORK IN THIS AREA

- The UK Data Archive has over forty years experience in selecting, ingesting, curating and providing access to social science data
- We have lots of experience of supporting researchers and data creators of social science data
- We support data sharing for the ESRC Data Policy (since 1995). We wrote the questions and guidance..!
- Our best practice approaches to making data shareable are based on:
  - challenges faced by hundreds of researchers to share data
  - handling others' social science data – quantitative and qualitative
- Highly skilled staff comprising researchers, technical and information specialists

PILOTING WITH ESSEX DEPARTMENTS

- Working on the ground to find out how departments are managing their research data
- Identifying both challenges and solutions that can be confronted and shared amongst other departments
- Identifying gaps, needs and weaknesses that the University can help with, e.g. long-term storage
- Can help support data management planning and meeting funders' requirements – e.g. RC grants

OUR PILOT DEPARTMENTS

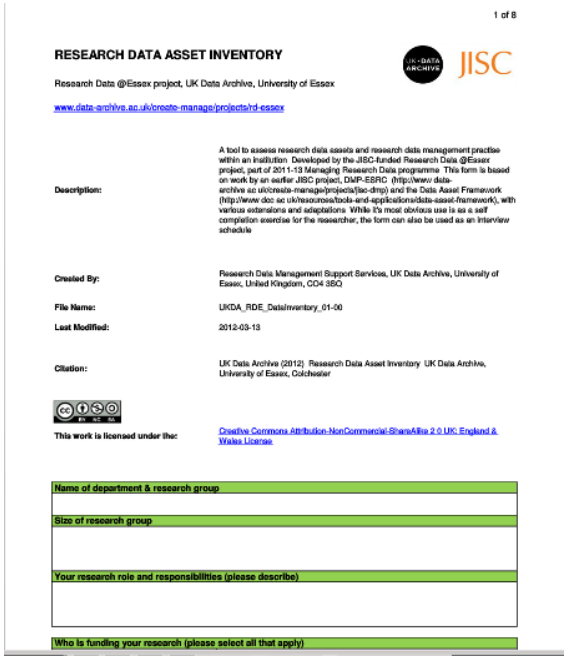
- Essex has a very broad research base for its size
- An institutional approach to research data management required careful assessment of varied research practise both between and within departments.
- Selected four divisions during our pilot assessment:
  - Department of Language and Linguistics
  - School of Biological Sciences
  - Essex Business School
  - School of Computer Science and Electronic Engineering
- Social science not chosen as UK Data Archive has a very good handle on these practices

DATA COLLECTION METHOD

- We spoke extensively to researchers on topics relating the management of research data, including data re-usability, access, storage and sharing.
- Pertinent research activities include everything from error checking the data as it is collected, to ensuring secure, backed up storage.
- Topics for discussion were chosen based on past data inventory work in the research data management community, particularly the Data Asset Framework.
- Our data inventory interviewing schedule borrowed from our past MRD work and other previous MRD approaches

RESEARCH DATA ASSET INVENTORY

- Schedule used to provide structure to the interviews
- Allowed respondents to deviate from the schedule, according to their own specific concerns and challenges
- Questions solicited suggestions as to what the University could provide
- We also had discussions with research leaders in pilot departments, who were able to provide a higher level perspective of incentives/sanctions to push forward the integration of data management



1 of 8

RESEARCH DATA ASSET INVENTORY

Research Data @Essex project, UK Data Archive, University of Essex

www.data-archive.ac.uk/create-manage/projects/rd-essex

Description:	A tool to assess research data assets and research data management practise within an institution. Developed by the JISC-funded Research Data @Essex project, part of 2011-13 Managing Research Data programme. This form is based on work by an earlier JISC project, DMP-ESRC (http://www.data-archive.ac.uk/create-manage/projects/jisc-dmp) and the Data Asset Framework (http://www.dcc.ac.uk/resources/tools-and-applications/data-asset-framework), with various extensions and adaptations. While it's most obvious use is as a self completion exercise for the researcher, the form can also be used as an interview schedule
Created By:	Research Data Management Support Services, UK Data Archive, University of Essex, United Kingdom, CO4 3SQ
File Name:	UKDA_RDE_DataInventory_01-00
Last Modified:	2012-03-13
Citation:	UK Data Archive (2012). Research Data Asset Inventory. UK Data Archive, University of Essex, Colchester
This work is licensed under the:	Creative Commons Attribution-NonCommercial-ShareAlike 2.0 UK: England & Wales License

Name of department & research group
Size of research group
Your research role and responsibilities (please describe)
Who is funding your research (please select all that apply)

GENERAL TYPES OF DATA

- **Qualitative**: collected through interview, recorded as audio (and sometimes video) and transcribed as text. Often have an associated annotation file associated with a transcript, which is highly valued by the researcher
- **Numerical, tabular**: typically handled in MS Excel (sometimes SPSS and other statistical packages), this kind of data forms the basis of a great deal of empirical research
- **Machine output**: logs or raw instrument-generated data. Often saved in proprietary formats, poorly documented, and hard to interpret without specialist knowledge
- **Cross-discipline collections**: emerging theme in a variety of disciplines. Typified by interdisciplinary research such as climate science, researchers producing these kinds of collection often collect the datasets together at a specific location

DATA SAMPLES COLLECTED

Biological Sciences

- Proteomics – mass spectrometry data from tumour tissue samples
- Bioimaging – high res image data collected to examine cellular structure

Business School

- Football managers –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

Computing and Electronic Engineering

- Artificial intelligence – crowd sourced AI scripts, and results from competition between these AIs in a game environment

TRIALLING A TEST-BED REPOSITORY

- Using Eprints platform – building on our existing Essex outputs repository

WE ARE BUSY RESEARCHING

- Features of new version of Eprints
- Metadata input
- Metadata streams to harvest from legacy repository
- Rendering of content
- Sword
- IR stats

- Metadata mapping to inform Eprints basic ingest metadata
  - Eprints default, DDI 2.1 &b 3.1, Inspire, Datacite, JISC Datashare
  - Watching the RepositoryNet+ space

EPRINTS VIRTUAL MEETING

- Meeting to discuss Eprints use for data
- Southampton team on hand
- Many synergies in thinking with IMDB team
- Discussion on problem of presenting the hierarchical relationships between studies, data collections and the multiple associated data and documentation files at the lowest level. Options:
  - follow the eCrystal type model and use a simple type classification for datasets within an Eprint
  - Use the Kultivate Containers plugin which tags a selection of EPrints with a shared type
  - Eprints team considering implementing a more integrated grouping for related items – but must consider knock on effects on the discovery/searching architecture and how to maintain a positive user experience

MULTIPLE METADATA LEVELS

- Need for multilevel metadata and to be able to drill down to different levels as necessary
- EPrints /IDMB thinking is 3 kinds of data within an Eprint of dataset-type:
  - Primary metadata
  - Additional metadata
  - Readme
- IMDB archaeology researchers - best way to approach metadata was to add an xml file containing anything that doesn't fit within the provided metadata schema
- Les Carr's IMDB 'Mind The Gap!' presentation - a small amount of clear, realistically acquirable metadata is better than a lot of rubbish metadata entered under duress!

ENTER THE DATA!

- Next step to adapt metadata fields in Eprints
- Complete the rendering on the page
- Ingest our sample data
- Complete metadata records (ourselves)
- Test and adapt search and browse functionality

- Demonstrate to depositors
- Discuss any access controls

- Showcase & feedback day on 19 June to participating departments and ISS, REO staff

e.g. Project or Series

Collection Title

Part of Container (If Applicable)

Name. *Title.* [Dataset] URI: http://isslx019.essex.ac.uk/id/eprint/###

Citation

Abstract

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Basic metadata upfront (blue underline = more-of-this link)

Keywords:	Lorem, ipsum, dolor
Division:	Department > Section
Subject:	H Subject 1 > Subject 1.1 F Subject 2 > Subject 2.1
Date Deposited:	22 Mar 2012 10:51
Depositing User:	Name
Last Modified:	22 Mar 2012 10:51

Detailed (but still generic) metadata on click-through

View details

Download

Download all files in one zip bundle

Download zip

All files (408k)

- Data files

Data.dat	104k	
SupplementaryData.dat	54k	
MoreData.dat	24k	

Item level metadata pop-outs OR tooltips

- Documentation

UserGuide.pdf	14k	
DataDictionary.pdf	54k	

Download individual files. Arranged by 'type' attribute

+ Discipline specific metadata

Discipline specific metadata included as item

Need to include description and/or extra metadata for each file? Tooltips?

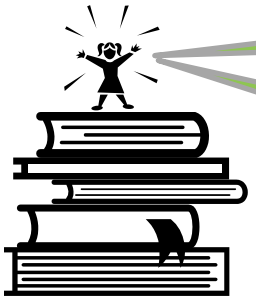
BUILDING DM SUPPORT INFRASTRUCTURE

- Established an informal Data Manager's Forum
  - First meeting 13 December 2011
  - Invited all Research Directors (who were keen & attended)
  - Considered not to be useful as a standing meeting
  - Research Directors quite keen to take on local information role and meet when critical matters arose
- Met with REO staff who support RC grants
  - Agreed placement of MRD support pages
  - Agreed possible centralised training – e.g. How to cost research proposals to include RDM and writing a DMP
- Have run one MRD generic course this year
- Successful training likely to be bespoke within faculties

BUILDING ON EXISTING LOCAL RESOURCES

- Managing and Sharing Data: best practice for researchers, the most recent version published in May 2011 (ESC and JISC grant)
- Managing and Sharing Data web portal and guidance for ESRC applicants
- Managing and Sharing Data: Training Resources pack – flexible training materials for people who train or support researchers and research support staff in how to look after research data (ESRC grant)
- Costing Data Management Guidance
- Data Management Strategies for Centres and Departments

SOME WORRIES AT ESSEX



I assume you back up all my data?

I keep my PhD interviews on this memory stick

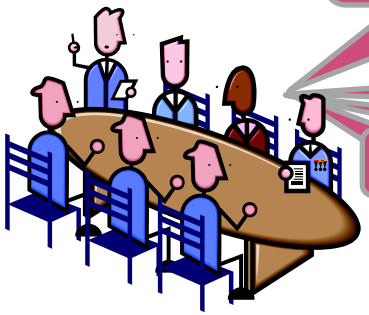
I have nowhere formal to offer my data

I have nowhere to store my large scale data

I'm worried about sharing my confidential data

Will Eprints allow me to share my data with my students but not others?

We don't really have a suitable collaborative environment



We use the cloud to collaborate

We need standards for data

We only use third party data sources – who owns IP on the derived data?

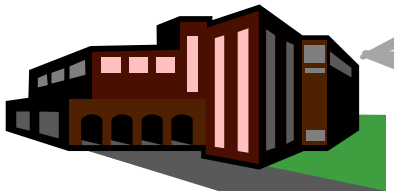
Our students could do with some training in this area

What happens to data when one of our investigators leaves?

We buy costly local data licences – could we do this at University level?

I am not sure if we collect any 'data'

Our resources are limited



Our priority is to meet the EPSRC Framework

Our policies must comply with the Data Protection Act

WE LOVE TO BLOG

- November 30, 2011 **Meeting the departments (and their data)**
- December 2, 2011 **Research data managment benefits and metrics: MRD 2011-13 Programme Launch Meeting**
- December 4, 2011 **Metadata session feedback: MRD 2011-13 Programme Launch Meeting**
- January 20, 2012 **Choosing a data repository platform: initial thoughts on Eprints**
- February 28, 2012 **Testing the Swordv2 PHP library with Eprints 3.3.8 on Debian Squeeze**
- March 20, 2012 **RD@Essex Progress Report (March 2012)**
- April 5, 2012 **EPrints for data: webinar report**
- May 2, 2012 **Research data asset inventory [download]**
- May 11, 2012 **DMP scoring system: RELU DMPs, 2005-2010**

THE RD@ESSEX TEAM

- Louise Corti - UK Data Archive
- Libby Bishop - UK Data Archive
- Tom Ensom - UK Data Archive
- Alexis Wolton – Information System Services
- Sarah Manning–Press – Research & Enterprise Office
- Sara Stock - Academic Section
- Bret Giddings - Information System Services

CONTACT

UK DATA ARCHIVE
UNIVERSIY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX CO4 3SQ

T: +44 (0)1206 872001
E: datasharing@data-archive.ac.uk
W: www.data-archive.ac.uk