University of Essex Staff Research Data Survey

A staff survey was held amongst all academic staff of the University of Essex, as part of the Research Data @ Essex project, carried out by the UK Data Archive in collaboration with the University of Essex Research & Enterprise Office and Information Systems Services.

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.

All university academic and research staff were invited to complete the online survey.

Essex Research Data is a pilot institutional data repository, developed from EPrints software. Traditionally software for publications repositories, we adapted EPrints by expanding the metadata profile for describing research data (based on DataCite, INSPIRE and DDI metadata standards) and by redesigning the data catalogue for presentation of complex collections.

Such data repository would allow institutional researchers to store and share data resources, and make them available to other researchers / users, either publicly or under restricted access conditions.
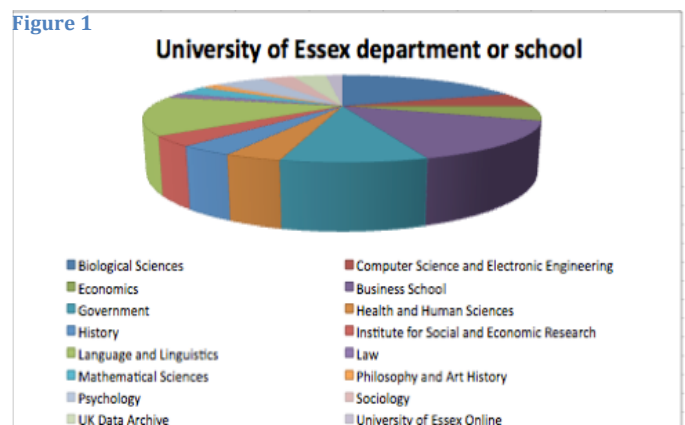
The survey was developed based on research data management expertise at the UK Data Archive, and also drew inspiration from a recent similar research data management survey held at Oxford University by the DaMaRO project.

The survey questionnaire and a summary report of all responses charts accompany this report.

Results

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%) (Fig 1). The 55 respondents represent 13


Figure 1

University of Essex department or school

- Biological Sciences
- Economics
- Government
- History
- Language and Linguistics
- Mathematical Sciences
- Psychology
- UK Data Archive
- Computer Science and Electronic Engineering
- Business School
- Health and Human Sciences
- Institute for Social and Economic Research
- Law
- Philosophy and Art History
- Sociology
- University of Essex Online

percent of all academic research staff at the university.

Respondents generate a mixture of data types, with numerical (70% of respondents) and textual data (60% of respondents) being the most prominent ones, besides images and audio data (24% of respondents each), multimedia data (16% of respondents), bibliographic data (12% of respondents) and geospatial data (4% of respondents)

Eighty percent of respondents produce relatively small data files of less than 2 gigabyte per file, and 20 percent generate 'big data' with file sizes over 2 gigabyte. During research, active data files are mostly stored on PC or laptop hard disk (80%), in online or cloud storage (51%) and on external hard disk (45 %). For longer term storage after projects end, researchers rely on hard disks, online and cloud storage and repositories (24%) (Fig 2). The main challenges they experiences with regards data storage are the sharing and storage of data in collaborative research, version control in collaborative research and storage capacity.

Figure 2 Short-term and long-term storage of research data
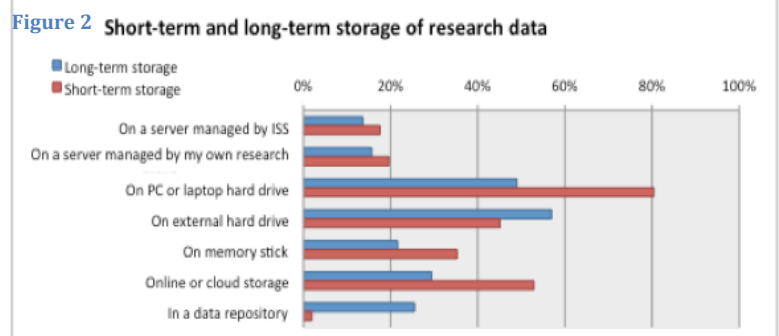
Eighty percent of respondents have in the last five years obtained external grant funding from the British Academy, UK research councils, the European Union, charities and private companies. Only a third of those obtaining external grant funding for research (30%) have currently been required to complete a data management and sharing plan as part of grant applications. Many (44%) are, however, aware that major research funders in their subject domain tend to require data management plans (Fig 3). 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.

Figure 3 Do major research funder(s) in your field expect information in funding proposals about how you will manage, preserve, and/or share created research data

Only twenty percent of researchers have received data management training or know where to access information. For most this is from the UK Data Archive.

Main data management training needs for respondents are (Fig 4):

- data management planning
- storage and backup
- ethical and legal aspects of data sharing
- costing data management, preservation and sharing
- Freedom of Information for research data
- copyright and intellectual property rights
- data formats, versioning and quality control for long-term preservation
- security and controlled access to data

Figure 4 Interest in topical support or training workshops

Overall, respondents are fairly familiar with and/or positive about sharing their research data. Forty-five percent of researchers have a funder requirement to share their research data. Twenty percent have legal requirements to do so. Twenty two percent have been required by a publisher to share or deposit research data and eighty one percent consider their data to be an important resource for future research and/or learning (Fig 5).

Figure 5
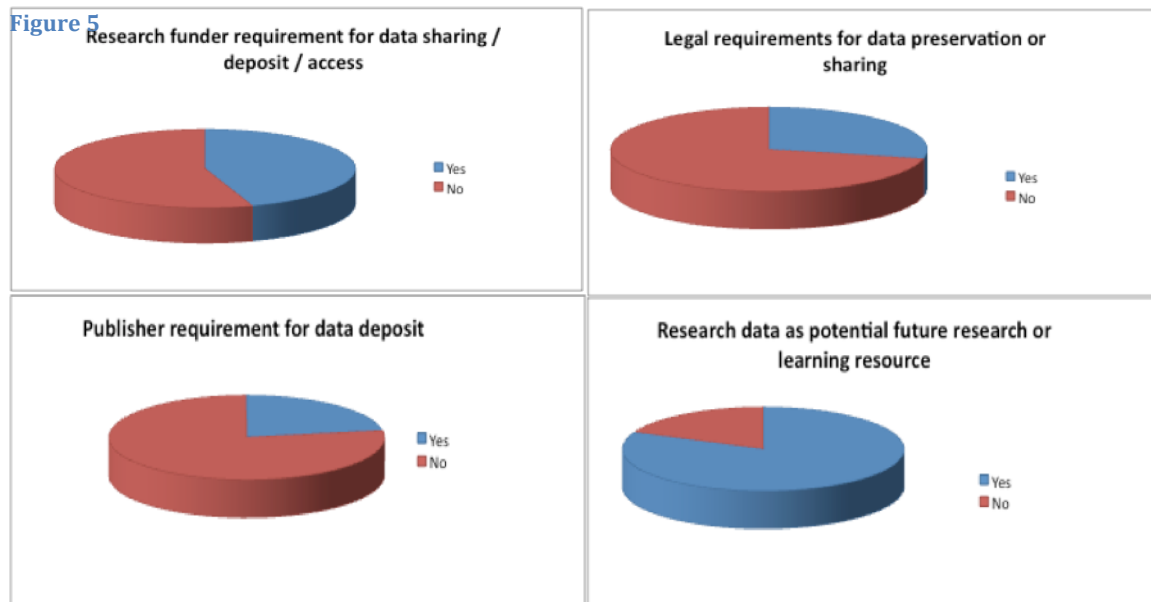
Forty five percent of researchers have placed data in a data repository, mainly because they are in favour of data sharing, because it is community practice in their research field, or due to a requirement from publishers or funders (Fig 6). The repositories or data centres used are:

- UK Data Archive
- British Oceanographic Data Centre
- Harvard IQSS Dataverse
- Interuniversity Consortium for Political and Social Research (ICPSR)
- Journal supplementary information or repository
- ArrayExpress
- British Athmospheric Data Centre
- Dryad
- Gene Express Omnibus (GEO)
- IRIS database
- Norwegian Social Science Data Services
- Pangaea
- Project website
- RCSB Protein Data Bank

Figure 6
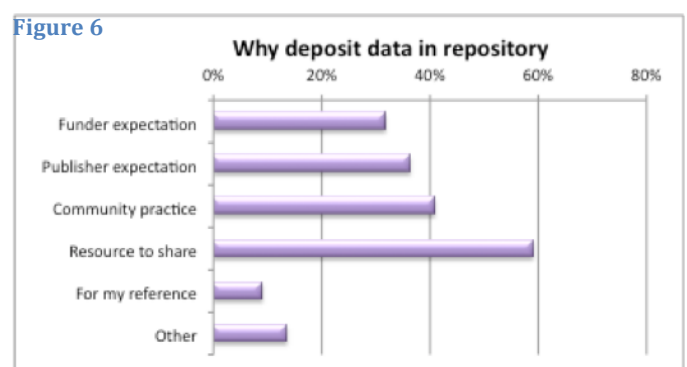
Those researchers not used to data sharing are either not aware of any suitable data repository or centre (13%), consider their data not suitable for sharing (16%) or do not find the time to prepare data for sharing (11%) (Fig 7).

Most respondents would consider sharing some of their data in future, either without restrictions, upon request or with known collaborators (Fig 8).

A quarter of respondents would like to deposit research data in a data repository at the University of Essex in future (Fig 9), but only 20 % are (very) likely to need to place data in a repository during 2013, for the following reasons:

- for data to be used by the research community
- to make data available
- for impact and visibility
- for research excellence purposes
- as funder requirement
- as journal requirements
- for longterm storage
- to promote collaboration

Figure 7

Why never deposited data in repository

Figure 8

Planned sharing of research data

- True for none of my data
- True for some of my data
- True for all or most of my data

Figure 9

Would you in future deposit research data in a data repository at the University of Essex?

- Yes
- No
- Don't know