RESEARCH DATA MANAGEMENT AT THE UNIVERSITY OF ESSEX

FINDINGS FROM A PILOT STUDY

THOMAS ENSOM AND LOUISE CORTI
RESEARCH DATA @ESSEX PROJECT
20 JULY 2012

T +44 (0)1206 874973
E tensom@essex.ac.uk
www.data-archive.ac.uk

UK DATA ARCHIVE
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX, CO4 3SQ

WE ARE SUPPORTED BY THE **UNIVERSITY OF ESSEX** AND THE **JOINT INFORMATION SYSTEMS COMMITTEE**

Contents

1. Background

The Research Data @Essex project is funded by the Joint Information Systems Committee (JISC) under the Managing Research Data programme 2011-13. The project was instigated to investigate current research data management practise at the University of Essex, and develop a best-practise strategy to adapt to evolving research data challenges. The project will encompass multiple disciplines and research groups in a whole-institution approach, and establish a test-bed institutional research data repository.

The UK Data Archive, based at the university, is taking a lead role on this project. The Archive's expertise in research data management, sharing and archiving will provide a crucial leg up for the university. It is imperative that the university prepares itself for stricter data management requirements from funding organisations, publishers and elsewhere in academia. Formalised data management policy is still very new to the university environment, although changes are accelerating as the UK research councils roll out policy statements. While the Economic and Social Research Council (ESRC) have had a Research Data Policy[1] in place since the mid-1990s, institutions were not asked to take any explicit role in data sharing. However, in 2011, the EPSRC took a lead in publishing a high-level policy statement[2] applicable to all funded institutions that wish to receive EPSRC funding, with compliance mandatory by 1st May 2015.

Our project has devised a multi-level strategy to tackle these institutional challenges, the basis of which is data inventory groundwork within individual departments to assess researchers' data management practices. The University of Essex has a very broad research base for its size, and as a result an institutional approach to research data management requires careful assessment of varied research practise both between and within departments. We selected and approached four divisions (either department or school) during our pilot assessment, in an effort to cover a broad range of research environments and data types:

- Department of Language and Linguistics (L&L)
- School of Biological Sciences (BIO)
- Essex Business School (EBS)
- School of Computer Science and Electronic Engineering (CSEE)

The UK Data Archive already has considerable experience with data management for social science disciplines, so our data inventory focused on less familiar research areas.

1.1. Definitions

Below are operational definitions of some of the key terms and concepts associated with the issues covered in this report. These definitions are taken or adapted from the UK Data Archive's Preservation Policy document[3] and 'Create and Manage Data' web resouces[4].

Data collection

A data collection is typically comprised of three components: data, documentation and metadata. Occasionally, a fourth component exists: code. Data collections are typically organised by reference to a particular survey or research topic and cover a specific geographic area and time period.

Data

Data are all the material, regardless of format, which are collected for reference and/or intended to be analysed. As part of datasets, they are the primary element of a data collection. More precise definitions of data vary according to context. Quantitative data may refer to just the matrices of numbers or words that comprise a data file, but may also cover other information (metadata) held within a statistical package data file, such as variable labels, code labels and missing value definitions. Qualitative data might include interview transcripts as well as audio and video recordings (analogue or digital).

Documentation

Documentation is that portion of a data collection that explains what the data mean. Documentation is required to understand data, and is therefore essential for potential re-use. It

commonly covers the subjects of sampling design, methods of data collection, questionnaire/interview design, structure of the data files, lists of variables and coding schemes, details of weighting, confidentiality and anonymisation, and provenance of any secondary data used. It also includes licence arrangements and all materials obtained through the original negotiation and data deposit, as well as post-deposit information created during preservation and ingest activities. The terms metadata and documentation are often used interchangeably and there is overlap between the two, though metadata are typically more structured and documentation tends to have a structure that is specific to each data collection.

Metadata

Metadata are a subset of core data documentation that are highly structured. Catalogue metadata provides standardised structured information explaining the purpose, origin, time references, geographic location, creator, access conditions and terms of use of a data collection. Metadata are typically used for resource discovery, providing searchable information that helps users to find existing data, as a bibliographic record for citation, or for online data browsing.

Data-level metadata can include file names and descriptions for variables, records and their values. They can also document codes and classification schemes used. Additional metadata for qualitative data can include document annotation such as textual mark–up.

Data management

Data management covers all aspects of handling, organising, documenting and enhancing research data. It is particularly important for facilitating data sharing, ensuring the sustainability and accessibility of data in the long-term, and allowing data to be re-used for future science.

Data management plan

For the effective management of data, planning starts when research is being designed and should consider both how data will be managed during the research and how they will be shared afterwards. This involves thinking critically of: how research data can be shared, what might limit or prohibit data sharing, and whether any steps can be taken to remove such limitations. Developing a data management plan should lead to the implementation of good data practices. Many research funders expect research data to be shared and therefore require data management and sharing plans before research starts. In the UK, the ESRC introduced a requirement for projects to complete a data management plan in 2011; also the Biotechnology and Biological Sciences Research Council (BBSRC) and the Wellcome Trust ask for data sharing plans to be a part of all research grant applications. In the US the National Institutes of Health (NIH) and the National Science Foundation (NSF) already require projects that collect data to submit data management plans as part of funding applications.

2. Methodology

We spoke extensively to research directors and researchers from pilot departments on the management of the research data assets they are responsible for. Research directors at Essex play a significant role in coordinating research activity within faculties. Their role includes dealing with: local data infrastructure needs, such as facilities and technical support; coordinating research funding and research proposals; staff and student project ethics review; research appointments and annual review; coordinating REF (Research Excellence Framework) returns and monitoring on-going research activity. These interviews focused on research data policy, infrastructure and other higher level issues and are covered in section 3.1 below.

Researchers interview were individuals who are actively collecting and managing a collection of research data. They may have a variety of roles outside of this, such as teaching or management of technical services for the department. In these interviews we covered the following topics, which form section 3.2 – 3.10 below:

- Data management strategy
- Documentation
- File formats
- Ethics and consent
- Copyright and IP
- Storage and backup
- File naming and versioning
- Sharing and re-use
- Security and destruction

Relevant research activities include everything from error checking the data as it is collected, to ensuring secure, backed up storage. The topics above were chosen based on past work within the research data management community. Creating a data inventory is a widely used method of assessing institutional research data assets, and we were able to draw from previous approaches by various research Council projects and Universities. As a basis for the structure of interviews we used our own interview schedule developed by UK Data Archive staff for the MRC research data management and sharing assessment in 2002[5] and further refined for the 2009 JISC DMP-ESRC project[6]. We also consulted JISC's Data Asset Framework[7]. The interview schedule was originally designed as a questionnaire for inventorying research centres so this necessitated some degree of revision and adaptation to suite a similar process across a whole institution.

This form was used as a schedule to provide structure to the interviews. An important part of our methodology was to also allow a respondent to deviate from the schedule, according to their own

specific concerns and challenges. Some questions were devised to give respondents a chance to suggest what they would like to see the university provide. As well as researcher interviews, we also had discussions with research directors in the pilot departments, who were able to provide a higher level perspective of the incentives and sanctions in place to push forward the integration of data management.

Each section below contains a summary of findings based on the interviews. It is our intention that these findings will feed into actionable recommendations and move forward infrastructure building and development of university policy.

3. Findings

3.1. Data management policy and awareness of local data management practices: research directors' perspectives

High level policy setting out requirements for management and sharing of local research data; existing local strategies and practices for managing research data

We questioned Research Directors about awareness of data policy in general and any specific gains or barriers they might envisage were a formal data policy to be put in place. Some had greater awareness than others of the types and location of research data in their faculties. Not all were aware of recent Research Council changes in data policy, but BBSRC, MRC and Wellcome Trust research -funded research departments showed greater awareness, largely through journals requiring data deposit alongside publication e.g. in the bio-sciences. Departments working with ESRC grants were also aware of data policy. The Centres receiving Engineering and Physical Sciences Research Council (EPSRC) grants were unaware of the new EPSRC requirements for universities, but were very interested to hear how their own university would be managing the strategy. The majority of research directors were very happy that the university supported them well in the grant submission process, but that longer-term data issues were not being questioned by research support staff. None were aware of any local problems with creating data management plans for Research Council proposals.

Most Research Directors interviewed agreed that short–term data management advice and infrastructure was handled quite well locally, but the longer-term was typically not planned for. In terms of data storage and longer-term legacy of data, all were concerned about the mass of data accumulating on department servers (including data acquired under license) and loss of control/orphaned data when a Principal Investigator (PI) leaves. Equally, in terms of data access,

most Research Directors felt that they were not informed about potential Freedom of Information (FOI) requests on data as they would like to be, and felt that university policy could help to provide much greater local awareness and procedures for responding to FOI requests. Many felt that there was no obvious centralised support from the university for these two areas – longer-term data storage and data access. A formalised data strategy at Essex would help to lay down rules and procedures for data storage and access, particularly about roles and responsibilities for governance and storage within the department and the University as a whole.

Barriers expressed by some, but not all, were that any formalised data sharing beyond their own like-minded research communities required by the university would present yet another load on top of already stressful research activity and may lead to problems of possible misuse of data. Many in the hard sciences already share data with colleagues via data registries or collaborative databases.

When asked about whether it would be useful to raise awareness of data policy and data management issues in their faculties, Research Directors agreed that their postgraduates would almost certainly benefit from advice on how to keep data safely and share appropriately. They did not any see any great need for their own staff to be trained, but induction on data governance issues like FOI and data protection, and on the university's recommendations for budgeting for research council bids regarding data storage, would be very useful. It was suggested that local data support people in departments, where they exist, could help take on this role. Those who knew about the UK Data Archive's role in data management training suggested that they could continue to run courses and awareness-raising events locally.

3.2. General data management strategy across departments

High level infrastructure and planning to enable the best practise management and sharing of research data

The following generic data types were found to be regularly collected in the university faculties we consulted:

- Qualitative: collected through interview, recorded as audio (and sometimes video) and transcribed as text. Often have an annotation file associated with a transcript, which is highly valued by the researcher.
- Numerical, tabular: typically handled in MS Excel (sometimes statistical analysis software such as SPSS), this kind of data forms the basis of a great deal of empirical research.

- Machine output: logs or raw instrument-generated data. Variable, but often saved in proprietary formats, poorly documented, and hard to interpret without specialist knowledge.
- Cross-discipline collections: these appear to be an emerging theme in a variety of disciplines. Typified by interdisciplinary research such as climate science, researchers producing these kinds of collection will often collect the datasets together at a specific location.

Data management planning (DMP) is a concept unfamiliar to most of the researchers interviewed . It seems this is not just a difference in terminology but a general lack of experience in formal planning for the long-term storage and sharing of research data prior to funding, such as completing Data Management Plans. As they become further integrated into the funding application process, demand for DMP guidance and tools will increase. Interviewees felt that the Research Enterprise Office (responsible for managing research project planning, proposals and grants) would be expected to take an active role in research data management planning and guidance. There was little awareness of the array of planning tools that already exist to suit different scenarios, such as those provided by the UK Data Archive[4] and JISC's Digital Curation Centre (DCC)[9].

One researcher we spoke to felt that there was too much variation in data management practise across the university and limited knowledge of best practise among staff. He felt that university support services providing a set of guidelines would help with problems of accountability and efficiency.

The majority of researchers interviewed expressed a strong interest and personal use case for a research data repository for the University of Essex. This interest was expressed not only in terms of sharing their own data, but also in getting access to others.

3.3. Documentation

Material explaining how data are created (and updated, where applicable), what they mean, and details of their content and structure

Most researchers are in the habit of keeping some form of documentation with their data, although this may be embedded in the data rather than a separate file. This documentation varies considerably with data type, but there are commonalities across cross-discipline within these types. Numerical data tends to be accompanied by variable information and keys to coding schemes. Interview transcripts and recordings are usually accompanied by a spreadsheet with information

about the respondents. More unusual examples of documentation-type materials are code and macros, which are sometimes used to automate the derivation of particular variables from new datasets. Collections methodology or derivation information tends not to be kept as a document with data, but will often be included in journal articles.

Unfortunately researchers tend to compile documentation with their own use in mind, rather than a totally new user. They may use idiosyncratic variable or file naming strategies using acronyms or non-transparent terminology. There is a danger that, without proper documentation, important details might be omitted and ultimately lost, impairing reusability.

3.4. File formats

Formats and software in which research data are stored and analysed

Most researchers are aware of open formats, and the importance of keeping copies of their data in those formats. However, other limitations currently necessitate the continued use of proprietary software. Among these limitations is the lack of availability and reduced analytic power of free software alternatives. A good example of this is found in biomedical imaging. Images are captured using highly specialised instruments, produced by an array of high-grade optics manufacturers. All of these instruments come with manufacturer-specific proprietary software, and thus file formats. Despite the option to export from the software into other formats, this often requires a compromise in not only convenience, but also the level of richness in the data and metadata.

Language and Linguistics researchers seemed to be the most familiar with the benefits of open formats, perhaps reflective of the ‘corpus’ culture of re-using datasets within sub-disciplines. Value added in this case comes from the addition of linguistic annotation files, again in open formats (typically XML or similar structured text formats). Business researchers on the other hand seemed the least concerned with open formats, as they very much viewed their data as for their own use and give little thought to reusability. Data services are commonly used to source third party data, so in many cases the data would be unsuitable for sharing anyway due to copyright issues.

For example, during the interviews we spoke to a research group in the biological science who are developing a mobile software application (app) that will be used to crowd-source data on aquaria from hobbyists. Archiving a mobile app as a research output could prove a challenge, and will necessitate the careful consideration of format and how best to maintain inter-file dependencies.

3.5. Ethics and consent

Those undertaking research to obtain informed consent for the re-use of data collected, and/or the implementation of procedures to avoid disclosure and thus protect the identity of individuals, organisations or businesses

Ethical issues most acutely affect those researchers dealing with human subjects – and we found that their awareness of ethical issues reflected this. Individuals undertaking research necessitating work with children were familiar with seeking approval from ethics committees (cross-discipline), as were those working with human tissue in biological research. Identifiable details such as names and postcodes were not found to be kept long-term. Intriguing emerging ethical challenges include neuro-psychological studies which use electrodes to monitor brain activity. In the future, there may be the potential for these studies to collect data on the current thoughts of a participant (although the likelihood of this is debated). This could obviously pose serious confidentiality issues and is likely to challenge existing notions of disclosive data.

Most, unfortunately, have not tried to use consent agreements allowing for their data to be shared. As a result, they felt that even anonymised transcripts from these data collections would not be releasable. Some had not considered that gaining consent for sharing at the time of fieldwork was an option. There is excellent guidance already available on how to do this, including information on the most appropriate wording to use.

3.6. Copyright and intellectual property

Clarity over the copyright and intellectual property rights of data

Intellectual property (IP) rights and copyright were a source of confusion for many. Some cited copyright of their research as lying with the university exclusively, while others, particularly those working in Biological Sciences, thought their funders would claim sole ownership as part of funding conditions. Perhaps this is due to more aggressive ownership assertion by discipline specific funders. Certainly, issues such as these need clarification if rights to sharing are to be in the hands of those generating data. Many researchers, particularly in the Business School, rely upon data streaming services (e.g. for daily stock market data). This is third party data for which the user holds only a licence to use, not complete ownership.

Desire to publish based on data was frequently cited as a reason to not release data – particularly as in many situations the same datasets might be expanded and re-used over a long periods of time (if not a whole research career). In these cases, IP rights should be critical in protecting first use and publication of findings. Equally prohibitive was the intellectual value of research data within commercial spheres. The proteomics researcher we interviewed in particular was concerned that his data might be of high value to pharmaceutical companies.

3.7. Storage and back-up

Strategy for sharing and storing research data, with rigorous back-up procedures to keep recent copies in case of original data loss

Storage issues were at the forefront of most researchers' minds when asked about data challenges they face. Many complained of a lack of clarity in terms of the storage facilities the university provides, and a further subset was unhappy about the limitations imposed on the space that was provided.

The challenge is a complex one, as storage requirements across the university seem to be extremely varied. Even within departments (see Figures 1 & 2) there is considerable variation. This is a product of massively varying volume requirements. Business researchers typically dealt with data that they could fit on their standard network storage allowance, and even then relied more on 'cloud' based data streaming services which could generate the required data on demand. On the other hand, research groups in Biological Sciences working with proteomic/genomic data could be generating hundreds of gigabytes of data per day. Some groups had bought their own server software which they either made accessible locally or occasionally through the university network to allow for wider access. There was general agreement that the university should improve the storage situation, particularly by offering secure and scalable solutions more tailored to a research environment.

Back-up procedures are currently largely dependent on whether they are automated by the storage solution, rather than any conscious effort by the user. Despite this, there were several examples of researchers keeping their own back-ups on physical media, and even sending copies of these to other parts of the world. At the other extreme, one biologist purchased and now maintains a multi-terabyte redundant (using RAID) storage solution for a very large collection of molecular data.

3.8. File naming and versioning

Procedures and checks to establish which version of a file is the most current, and record update history

Naming conventions are not consistent between researchers, but are sometimes maintained within a particular group. Despite a lack of consistency, this issue does tend to be handled quite well within research groups so as to avoid confusion over file contents and relationships. Typically sampled entities are given identification numbers or codes through which they are referred to in data.

Versioning however is typically managed in a much more haphazard way by researchers, and was an issue raised consistently in interviews. Problems seem to stem primarily from a lack of suitable file management software, which becomes necessary when dealing with large numbers of files and complex relationships. Versioning can be a particular problem when there is extensive processing or derivation of additional variables required after data collection (e.g. a survey dataset), these processes potentially requiring multiple copies be produced and tracked. At the other extreme, one researcher in the Business School told us that it would be easier to re-download and analyse data rather than storing it and managing versioning.

There is a significant divergence in the regularity with which updating is required. Some researchers require datasets to be updated on only a yearly basis (e.g. data derived from finance annual reporting) while others are carrying out minor additions on a daily basis (e.g. genomics). As a result, there is no blanket solution for version control across departments; this challenge will require individual strategies to be developed. Standard file locking (i.e. single user) methods are not necessarily practical in a modern instrument-based research environment where files have the potential to be very regularly updated. The use of Virtual Research Environment's has great potential for data management in general, and for versioning in particular; by integrating and automating control of this so it fits seamlessly into normal workflows.

3.9. Sharing and re-use

Plan for sharing and future use of data, including consideration of funder requirements

Sharing within research groups is common practise, and many researchers expressed an interest

in more sophisticated ways of doing this. Sharing is required not only within the university, but also externally with collaborators at other institutions. However, the majority of researchers did have reasons preventing them from wanting to openly share at least a subset of their data. Some researchers appreciated the potential for other departments to hold data of potential value to them, and view the project as a way of getting at this (accurately or otherwise). These attitudes are pervasive in research communities, and include:

- A fear of plagiarism before publication of research
- Complete consent for sharing might not have been granted by research subjects
- The belief that their data is of no interest to others

While these fears are understandable, most are ameliorable or insignificant in contrast to the many potential benefits of data sharing. Plagiarism risk can be mitigated by putting a prepublication embargo on a dataset. If the original consent form did not obtain permission for data sharing, retrospective consent can sometimes be obtained. If consent has not been explicitly denied then there may still be the option to share anonymised data. Such situations should be dealt with on a case-by-case basis. There are many cases of data having unexpected re-use value in research, and this is particularly significant given the impossibility of predicting future research challenges. There may be problems though, with data acquired in uncertain terms. One researcher mentioned “gentlemen’s agreements”, ambiguous ownership terms and “who you know” driven exchanges in his field. These are indicative of institutionalised limitations to data sharing based on perceived ownership and value, and reinforce the need for early and comprehensive data management planning.

The desire for time in which to work on publications based on data was frequently mentioned as preventing the more immediate release of data – particularly as in many situations the same datasets might be built-on and re-analysed over long periods of time. The other significant obstacle for researchers is fear over commercial value of datasets, for example the researcher mentioned in section 3.6 who had concerns over the value of his biomedical data to pharmaceutical companies.

There were reservations among some about the use of external archiving services. They are distrusted in certain fields (particularly biomedical) due to being perceived as unstable and untrustworthy. One researcher offered anecdotal evidence of past archives having disappeared (along with their data holdings) due to funding cessation, and that there have been cases of archives selling on data entrusted to their care. It is important that the legacy of this history be addressed in the advocacy of stable archiving solutions.

3.10. Security and destruction

Secure methods for the storage and transmission of data files, finding alternative arrangements with secure data facilities where necessary. Also defined procedures for the secure erasing of data when required

Security was a concern of only a subset of those interviewed. Primarily, this challenge manifested itself where data were obviously sensitive – for example, medical research involving human tissue, or the personal details of interview respondents (name, address, date of birth). Despite showing concern, data security did not seem to be a high priority for these researchers. Although breaches do seem to be very rare events, there was one striking example. A concerned research facility manager told us of that their central storage unit had recently been remotely hijacked and used to store copyright infringing material.

Data seems to rarely be transmitted in an encrypted state when transferred among collaborators. Methods of transfer encountered include:

- Posting physical copies as parcels
- Use of file-transfer services including the Universities own, ZendTo, as well as cloud based (e.g. Dropbox)
- Use of shared university network drives

A lack of thought over long-term storage of data was generally apparent among those we spoke to. Data retention rarely seems to have been formally agreed with funder or institution. Medical research was the main exception, with limited retention periods sometimes agreed as part of the ethical approval process. In some cases informal retention periods existed, mainly due to insufficient resources to maintain an exponentially increasing volume of data in the long term. These informal retention periods were not strictly adhered too. Certain research areas, for example Language and Linguistics, would not consider retention a relevant issue, as their work relies on building up a database of texts (a ‘corpus’) that a researcher may work with for their entire career.

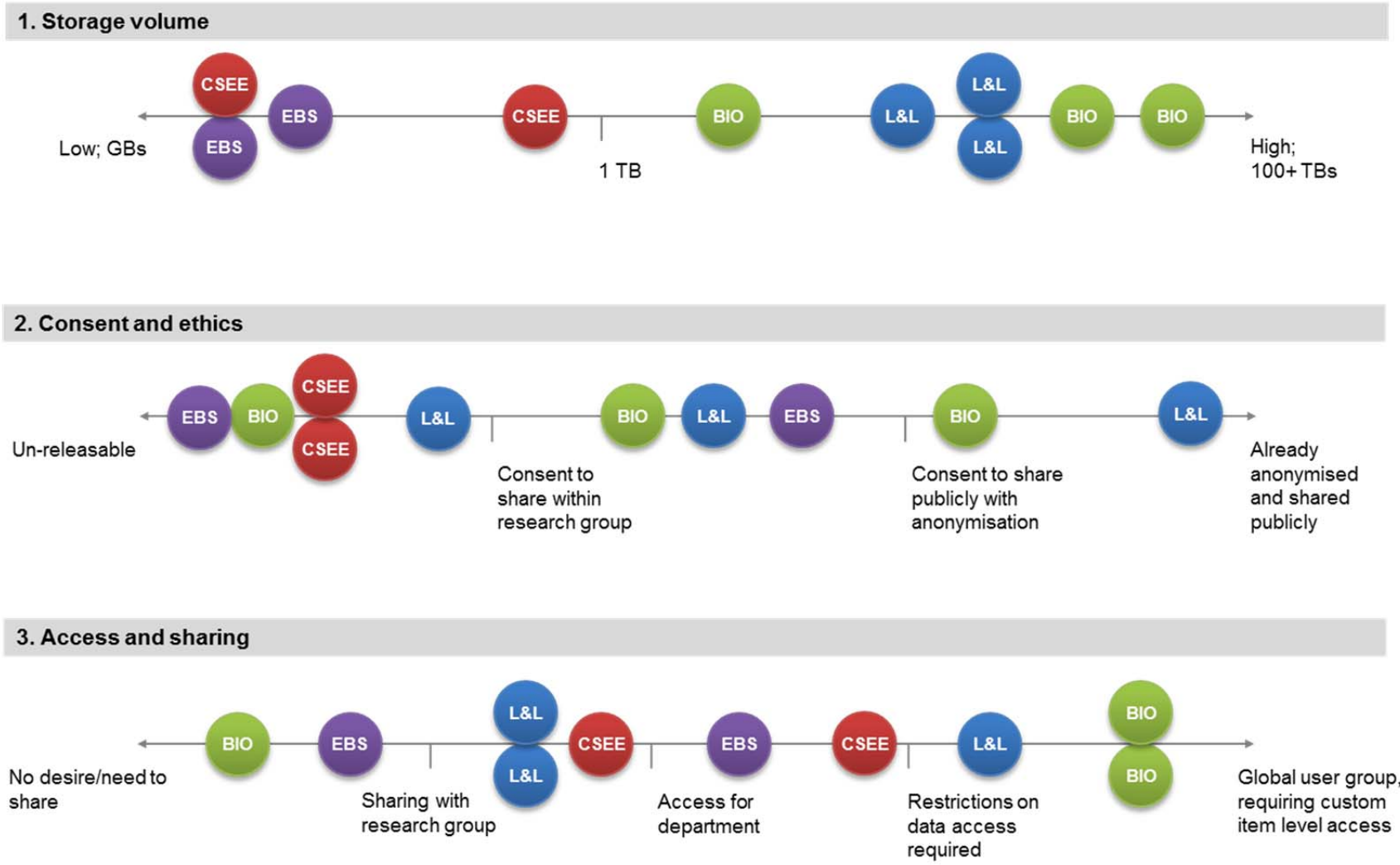

Figure 1. Visualisation of data management challenges in inventoried departments: Department of Language and Linguistics (L&L), School of Biological Sciences (BIO), Essex Business School (EBS), School of Computer Science and Electronic Engineering (CSEE). Each circle represents a generic data collection for a particular researcher. The position of the circle horizontally indicates where along the spectrum for that particular challenge this typical dataset lies.

4. Summary and concluding remarks

The findings and recommendations covered in details in section 3 of this report are summarised in Table 1 below.

Table 1. Summary of findings on research data management at the University of Essex, based on interviews carried out by the Research Data @Essex project

TOPIC	FINDINGS
DATA MANAGEMENT POLICY & AWARNESS OF LOCAL DATA MANAGEMENT PRACTICES *High level policy setting out requirements for management and sharing of local research data; existing local strategies and practices for managing research data*	• Local control and management of data, varying by department and research group • Little awareness of any formal roles and responsibilities for governance and storage within departments or the University as a whole • No centralised support from the University for longer-term data storage and data access. . • Little awareness of any structures and procedures for responding to FOI requests for data, or of forthcoming ESPRC Policy • No real thought given to costing longer-term data storage and access • Absence of training for staff and post-graduate students on data management and data storage and security issues
DATA MANAGEMENT STRATEGY *High level infrastructure and planning to best enable the management and sharing of research data*	• Common generic data types across all departments and disciplines • Data management planning is a new concept to many researchers, but is likely to soon become an integrated part of the research funding process • A lack of centralised guidance on research data management issues • A research data repository would be valued
DOCUMENTATION *Material explaining how data are created (and updated where applicable), what they mean, and details of their content and structure*	• Researchers do tend to keep documentation with their files, although this often with their own use in mind • Documentation materials may not be conventional (e.g. text files) but could include code and macros

FILE FORMATS *Formats and software in which research data are stored and analysed*	• Many researchers use open formats to store their data already, but are ultimately limited by available instruments and analysis software • Conversion from proprietary to open format can result in the loss of quality, metadata or detail
ETHICS AND CONSENT *Those undertaking research to obtain informed consent for the re-use of data collected, and/or the implementation of procedures to avoid disclosure and thus protect the identity of individuals, organisations or businesses*	• Much research at the university involves some kind of ethical consideration, and knowledge of procedures seem to generally reflect this • Most consent agreements were found to not take into account data sharing
COPYRIGHT AND IP *Clarity over the copyright and intellectual property rights of data*	• Understanding of copyright was found to generally be confused among researchers • Assertions of ownership seem to vary depending on funder, even within UK Research Councils
STORAGE AND BACKUP *Strategy for sharing and storing research data, with rigorous back-up procedures to keep recent copies in case of original data loss*	• Perceived lack of clarity in the universities provision of storage for research data • Wildly varying requirements in terms of storage volume • Backup practises are variable, and sometimes based on limited knowledge of the available options
FILE NAMING AND VERSIONING *Procedures and checks to establish which version of a file is the most current, and record update history*	• File naming and versioning are typically handled pragmatically, but are often time intensive • Some frustration over lack of solutions to these two issues is apparent
SHARING AND RE-USE *Plan for sharing and future use of data, including consideration of funder requirements*	• General attitudes towards data sharing are negative, which is limiting movement in this area • Key issues include concern over risk of plagiarism , and disclosure

SECURITY AND DESTRUCTION *Secure methods for the storage and transmission of data files, finding alternative arrangements with secure data facilities where necessary* *Also defined procedures for the secure erasing of data when required*	• Sensitive data is routinely handled at the university, and but security measures have not always been sufficiently considered • Retention period of research data is not usually formalised

The University of Essex is well positioned to take a leading role in the development of comprehensive research data management, planning and sharing. Particularly beneficial will be the presence of the UK Data Archive on campus as an advice and support base, who are internationally known for their work in promoting and supporting data sharing and training in research data management. The continued monitoring of the research data management climate and community activity will ensure not only that new challenges are swiftly dealt with, but also that the latest innovations are at the disposal of the institution. It is essential that the University of Essex works to address the challenges laid out in this report if it is to enhance its reputation as a forward thinking academic institution with a strong research base.

In terms of wider reach, our findings provide valuable information on the state of research data management in a multi-discipline, research focused institution. It is our hope that in publishing this report, that these findings might be of interest and of value to similar initiatives.

5. References

1. ESRC (2010) ESRC Research Data Policy http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf

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

3. UK Data Archive (2011) UK Data Archive Preservation Policy. 18 May 2011, http://www.data-archive.ac.uk/curate/preservation-policy

4. UK Data Archive (2012) Create and Manage Data http://www.data-archive.ac.uk/create-manage

5. Corti, L. & Wright, M. (2002) MRC Population Data Archiving and Access Project: Consultants Report. http://www.data-archive.ac.uk/media/1690/MRC_UKDataArchive_DSP_consultantsreport.pdf

6. UK Data Archive (2012) Essex Research Data Inventory.

http://researchdataessex.posterous.com/110349327

7. Data Asset Framework (2009) Implementation Guide. October 2009, http://www.data-audit.eu/

8. Horton, L., Van den Eynden, V., Corti, L. and Bishop, L. (2011) Data Management Recommendations for Research Centres and Programmes. UK Data Archive, Colchester: 28 March 2011

9. Digital Curation Centre (2012) DMPonline https://dmponline.dcc.ac.uk/

Table 1. Departments, research areas and typical datasets from the individuals interviewed during the project.

Department	Area	Typical dataset
Biological Sciences	Proteomics	Mass spectrometry data from cancer tumour tissue samples
	Ecology	Interdisciplinary studies of coral reefs, including biodiversity surveys, images and environmental analyses
	Sports Science	Bio-social survey of young people's exercise habits
	Bioimaging	High resolution images of biological material at a microscopic level (e.g. cells)
Essex Business School	Management	Empirical data on managerial performance and succession, used to examine performance metrics
	Finance	Daily stock market data delivered through paid subscription service
Language & 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
	First language acquisition	Audio and transcripts of interviews with multiple generations of Indian English speakers
Computing and Electronic Engineering	Artificial intelligence	Crowd sourced AI scripts and the results of competition between these programs
	Brain-computer interfaces	A combination of neuro-physiological measurements, machine logs and computer programs