IJDC | *Peer-Reviewed Paper*

Research Data Management Practices: A Snapshot in Time

Mary Anne Kennan
Charles Sturt University – Sydney

Lina Markauskaite
University of Sydney

Abstract

There is increasing pressure from funders, publishers, the public, universities and other research organisations for researchers to improve their data management and sharing practices. However, little is known about researchers' data management and sharing practices and concerns. The research reported in this paper seeks to address this by providing insight into the research data management and sharing practices of academics at ten universities in New South Wales, Australia. Empirical data was taken from a survey to which 760 academics responded, with 634 completing at least one section. Results showed that at the time of the survey there were a wide variety of research data in use, including analogue data, and that the challenges researchers faced in managing their data included finding safe and secure storage, particularly after project completion, but also during projects when data are used (and thus stored) on a wide variety of less-than-optimal temporary devices. Data sharing was not widely practiced and only a relatively small proportion of researchers had a research data management plan. Since the survey was completed much has changed: capacities and communities are being built around data management and sharing and policies, and guidelines are being constructed. Data storage and curation services are now more freely available. It will be interesting to observe how the findings of future studies compare with those reported here.

Received 30 September 2014 | *Revision received* 10 June 2015 | *Accepted* 1 July 2015

Correspondence should be addressed to Mary Anne Kennan, Charles Sturt University, Locked Bag 450, Silverwater NSW 2128 Australia. Email: mkennan@csu.edu.au or Luna Markauskaite, CoCo Research Centre, University of Sydney, NSW 2006, Australia. Email: lina.markauskaite@sydney.edu.au

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction and Background

> "Because I am no longer storing my data on the Arts Faculty shared server, I feel as though I am dicing with death."

During the last decade, national and international attention has been increasingly focused on promoting research data management and access to publicly funded research data. Researchers increasingly feel pressure to improve and sometimes radically change their data management and sharing practices. This pressure comes from research funders seeking to add value to expensive research and solve cross-disciplinary grand challenges, from publishers seeking to be responsive to calls for transparency and reproducibility of the scientific record, organisations seeking to manage potentially valuable data, and the public seeking to gain and reuse knowledge for their own purposes using new online tools (Borgman, 2007).

However, researchers' perspectives about data management questions and issues have been little understood. Lack of awareness of researchers' needs and limited engagement with researchers in co-development of policies and infrastructures may not only disenfranchise many researchers, but may also actively discourage short- and long-term uptake good data management practices.

This paper provides some insights into research data management practices of university academics in New South Wales in Australia in 2011 and 2012. It builds on survey responses about eResearch practices collected from academics in ten institutions. In this paper we have three aims: a) to shed some light into researchers' data handling and management practices, needs and constraints at a particular point in time; b) to discuss them in the light of data management initiatives and debates; and, by doing this, c) to contribute to the discussion about the importance of understanding these practices in order to inform the development of data management polices and infrastructures. The paper concludes with suggestions for further studies for which this paper could provide interesting comparisons. It would be interesting to investigate how data management practices have developed since 2011-12 and to what extent national support initiatives, such the Australian National Data Service (ANDS) and the relatively new requirement (for applicants from 2014 onwards) for the inclusion of data management plans in the national grants programs of the Australian Research Council (ARC) (Australian Research Council, 2013, 2014a) have made an impact. In late 2014 the other major Australian research funder, the National Health and Medical Research Council (NHMRC), issued a targeted consultation draft of *Principles for Accessing and Using Publicly-funded Data for Health Research* (NHMRC, 2014) which seeks to maximise research use of publicly funded health and health-related data.

At the time the survey reported in this paper was conducted there were no national requirements for researchers to have data management plans. However, the ARC's Discovery Projects Funding Rules for funding in 2015 noted that "the ARC considers data management planning an important part of the responsible conduct of research and strongly encourages the depositing of data arising from a Project in an appropriate publicly accessible subject and/or institutional repository" (Australian Research Council, 2014b); and the continuing implementation of institutional research data management policies and supporting infrastructures may have improved the situation (Beitz, Groenewegen, Harboe-Ree, Macmillan, and Searle, 2013; Treloar, Choudhury, and Michener, 2012).

In order to address these aims, we first briefly address the literature relevant to this study and provide an explanation of the research method and then present our findings about data archiving and preservation from a researchers' perspective. We conclude with the recommendations that in addition to provision of infrastructure (such as storage) and services (such as curation) there are also issues for researchers about which further discussion and/or education is necessary. These include issues identified by researchers such as how to manage ethical concerns, privacy and confidentiality; how to make decisions about the future usefulness of data; and how to provide support with technical issues, such as metadata creation and curation across the lifecycle to prevent obsolescence and loss.

Literature

Nature of Data

The main focus of this paper is data collected for research purposes in universities. Research data are defined by Rice (2009) as data "collected, observed or created for the purposes of analysing to produce original research results." However, increasingly Rice's definition may be too narrow, as data used in research may be "collected, observed or created" for purposes other than research, and then later used in research. Examples include the administrative records, log files of learning management systems and web portals and other behavioural traces used in learning analytics (Charlton, Mavrikis and Katsifli, 2013; Scifleet, Henninger and Albright, 2013) or the traces of individual lives available from social media and online discussions (Godbold, 2013; Hey, Tansley and Tolle, 2009; Verbert, Manouselis, Drachsler and Duval, 2012). Research data are heterogeneous because they can take many forms depending on their origins, the research problem being addressed and the discipline of the researcher. Data sources are also varied. In the life and physical sciences, data are generally gathered or produced by researchers through observations, experiments or by computer modelling. Borgman (2007) provides some examples of data from different scientific disciplines: X-rays in medicine, protein structures in chemistry, spectral surveys in astronomy, specimens in biology, and events and objects in physics. In the social sciences researchers may gather or produce their own data from, for example, interviews, surveys and questionnaires, and observations. Research data can also be obtained from third parties, for example the Australian Bureau of Statistics or the Organisation for Economic Co-operation and Development. Humanities data often come from cultural records, archives and objects; both published and unpublished (Borgman, 2007).

Researchers have always created and collected data, but, with emergence and growing availability of digital data and global computer networks, the resulting increases in data set sizes and complexity create new challenges for use, reuse, storage, future access and retrieval. The increasing use of computer-based data collection, conversion and capture instruments and sensors has expanded the scale of research data in many disciplines. Researchers are thus confronted with data overload in the form of vast collections and arrays of data that have been generated from vast surveys of, among other things, the galaxies, physical phenomena on earth, and the molecular composition of organic life, and human behaviour in digital space. In contrast to data from traditional controlled experiments or applications of theory, these data are not easily usable for analysis and theory testing, and require new kinds of research skills and practices. Some

 doi:10.2218/ijdc.v10i2.329

have called for the development of what has been coined “data literacy” (Mandinach and Gummer, 2013) which has even been formally incorporated in the teaching of educators (Carlson, Johnston, Westra and Nichols, 2013). As some researchers have suggested, a new paradigm called “data-driven scholarship” is emerging or what Grey calls “data-intensive science … the fourth paradigm” (Grey, 2009).

Issues specific to digital data management and curation include:

- Technology obsolescence;
- Technology fragility (e.g. corruption of files);
- Lack of guidelines on good practice;
- Inadequate financial and human resources to manage data well;
- Lack of evidence about best infrastructures (Harvey, 2010).

The activity of managing and promoting the use of data from their point of creation or collection, ensuring they are fit for contemporary purpose, and available for discovery and later sharing or re-use is “data curation” (Harvey, 2010). For dynamic data sets this may mean continuous enrichment or updating to keep it fit for purpose. Harvey also adds that curation holds the promise of ubiquitous access, deals with the fragility of data ensuring its long term accessibility, preservation, authentic and integrity, manages risk, the obsolescence of hardware and software, addresses the need for well-constructed file systems and metadata, and manages intellectual property rights. Another key issue in data sharing and reuse is the issue of provenance (information about where and how the data originated, how it has been stored, indexed, modified or used), which is usually managed via annotations in metadata descriptions. Information about data provenance needs to be updated when any change or use is made of data throughout the entire lifecycle of the data and its accuracy is critical in research (Andersson and Sørvik, 2013; Harvey, 2010). Data provenance becomes one of key enablers for successful evolution of more open, accountable and collaborative data-intensive research (Bechhofer et al., 2013; Glavic, 2014; Michener and Jones, 2012).

Obtaining and Sharing Data

Until recently data have rarely been seen by people beyond the initial research team. Data are usually analysed, summarised and published, often after being theorised, as text in articles and books. Analysed and published data are generally selective representations of a small amount of the raw data originally collected by the researchers (Latour, 1987). Publications, such as this, inevitably incorporate methodological and pragmatic choices made by the researchers at different stages of the research and limit subsequent interrogation of the data. Data that are abstracted and prepared for publishing in these ways do not necessarily provide sufficient information for future users (Markauskaite, 2010). These considerations have led to the rise of the Open Data (OD) movement (Pampel and Dallmeier-Tiessen, 2014). Open Data is a term defining how scientific data may be published and reused without price or permission barriers. OD is related, but not completely analogous, to open access (OA) (Murray-Rust, 2008).

Drawing on scholars who have been investigating these issues we could summarise that even if data are not to be completely openly stored and shared, it is important to store, describe and provide the possibility of sharing, because data:

1. Are expensive to collect and therefore publicly funded research should be publicly available (Murray-Rust, 2008);
2. May be unique and impossible to replicate, such as data representing a snapshot in time or space (Henty, Weaver, Bradbury and Porter, 2008);
3. Can be reused to reproduce and validate original findings, to advance the original research or to open another line of enquiry (Witt, 2009);
4. Can contribute to answering questions that may require inter-disciplinary problem solving (Cragin, Palmer, Carlson and Witt, 2010);
5. May be used to examine a phenomenon from different epistemic or social perspectives (Markauskaite, 2010);
6. May need to be collected and integrated from a variety of sources, beyond the scope of one research team, time or location (Borgman, 2007).

Several of the points above imply the need for data sharing, which is a key element of collaboration (Borgman, 2006). Altruism and the potential for new collaboration opportunities may motivate some researchers to share their data but, as studies show, this goal is low on the list of researchers' priorities (Henty et al., 2008; Markauskaite, Kennan, Richardson, Aditomo and Hellmers, 2012).

Data Storage, Preservation and Management

The ongoing storage, preservation and curation of data is increasingly important given the potential for reuse and sharing of data in new ways as discussed in the section above. While researchers tend to use the term "storage" when referring to what happens to data not actively in use, there are a number of technical terms that are more specific and relevant. "Archiving" for example, is a curation activity which ensures that data is properly selected, stored, can be accessed and that its logical and physical integrity is maintained over time. Archiving includes maintaining security and authenticity.

"Preservation" is an archiving activity in which specific data are maintained over time so that they can still be accessed and understood through successive change and obsolescence of technologies, such as the hardware on which data are stored and the software by which data are accessed. Preservation activities often also include policies, strategies and procedures (Harvey, 2010).

The useful life of data and data collections is referred to as the data lifecycle. Decisions made at each stage of the data lifecycle determine what data is available at the next stage, how it is handled, and the purposes for which it is useful. The lifecycle of data will be different in different disciplines or research traditions. At different stages in the lifecycle, different people will be responsible for the data. There are many different data lifecycle models, including that of the Digital Curation Centre (DCC) in the United Kingdom, which focuses on data curation and preservation issues (DCC, n.d.; Higgins, 2008; Higgins, 2012) and that of the Data Documentation Initiative (DDI, n.d.) both of which elaborate researcher and curator roles in data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving.

Unmanaged data present financial and opportunity loss, and yet the description, long term storage and preservation of data are not necessarily key concerns of researchers who create, collect and use data. Effective data management must begin with the start of the project and should be done by the researcher or research team (Pryor, 2012). Good data management requires input from a team of stakeholders: from the research team, to university policy makers, information technology and information management staff, and librarians or information managers.

Researchers, research funders and research institutions, such as universities, place increasing importance on data management planning as a way of improving the access to, and the longevity, sharing and reuse of research data (Pryor, 2012). Formal research data management is a developing area; practices are emerging. Data management has traditionally been an orphan' activity expected to be carried out, but not explicitly taught, supported or funded (Donnelly, 2012). Formal research data management planning, in most disciplines, is a relatively new activity. However, as Donnelly points out, neglecting to make data management and planning a formal requirement is risky for institutions and funders, as activities which do not attract funding, reward or recognition are the first to suffer when time or money get tight. Thus it is important for data management plans to have associated policies and systems to support researchers in their implementation.

Decisions made in relation to the management of research data should be informed by relevant legislation and codes; national and institutional policy; and procedures and guidelines that the research project must adhere to. These will vary in different countries and states. The examples given here are largely Australian, but similar ones will inform research practice in many other countries. A plan enables compliance issues to be visible and achievable. For example, in New South Wales, institutions and researchers will be guided by the provisions of the State Records Act 1998[1] regarding the retention and disposal of research data. Various copyright acts may also apply to data, including to digital forms of text and images, packaged variously as videos or DVDs, for example. The code of computer programs (both the human readable source code and the machine readable object code) is protected by copyright as a literary work. Data compilations such as data sets and databases can also be protected by copyright. Although, as a general rule of copyright, ownership resides with the author or creator, copyright can belong to a person's employer or via contract to others, depending on conditions of employment or via contracts in multi-institutional collaborations. If researchers wish to share their data by depositing into a database, repository or archive, share their data online, or allow others to access and use their data in any way, they will need to build appropriate permissions and licenses into their data management plan and metadata.

In Australia, the Australian Code for Responsible Conduct of Research guides institutions and researchers in responsible research practice. The Code has been jointly developed by the National Health and Medical Research Council (NHMRC), the Australian Research Council (ARC) and Universities Australia. Compliance with the Code is a pre-requisite for NHMRC and ARC funding. A central aim of the Australian Code for the Responsible Conduct of Research is that sufficient data and materials are retained to justify the outcomes of research and to defend such outcomes should they be challenged. The Code also makes recommendations about what research data and associated information should be kept (NHMRC, ARC, & Universities Australia, 2007). Similarly, national ethical codes (NHMRC, ARC, & AVCC, 2007) make recommendations about the ethical considerations that should be considered by researchers regarding their data. A data management plan enables processes and checks to be put in place to ensure such requirements are met.

Increasingly funders and institutions are requiring researchers complete data management plans and have policies in place to ensure that research data meets these requirements (Borgman, 2012; Corrall, Kennan and Afzal, 2013; Groenewegen and Treloar, 2013). The ARC and the NHMRC did not require data management plans at the

1 NSW States Records Act: http://www.records.nsw.gov.au/about-us/state-records-act-1998

time of the survey, however from 2014 a section of grant application forms titled "The Project Description" requires researchers to articulate briefly their plans for the management of data generated through the proposed Project. Nor has there been a mandate for researchers to deposit research data in repositories, although the ARC strongly encourages this (Australian Research Council, 2014b).

Over the years there have been small scale studies of academic research practices, including data management and sharing practices, in Australia (Henty et al., 2008; Houghton, Steele and Henty, 2004; Markauskaite, Hellmers, Kennan and Richardson, 2009). These studies essentially indicate that many researchers do not have formal data management plans, and have concerns about the formal management of their data which they generally have not been in the practice of sharing. A larger study based in the United States found similar results and that barriers to effective data management, sharing and preservation were rooted in the practices and culture of the research process (Tenopir et al., 2011). New mandates for data management plans from the National Science Foundation and other federal agencies in the US and by the Research Councils in the United Kingdom, for example, bring attention to the need to manage, share and preserve data and could lead to changes (Tenopir et al., 2011; Wallis, Borgman, Mayernik and Pepe, 2008), as could support offered by organisations such as the Digital Curation Centre (DCC)[2] in the UK and the Australian National Data Service (ANDS)[3]. This study adds to our knowledge of existing academic practice in this area and will provide useful benchmarking information for future studies.

Method

The study was conducted in late 2011 and early 2012, using an online survey which asked questions about three main eResearch areas: a) data management, retention and sharing; b) technology-enhanced research methods, tools and services; and c) research collaboration and dissemination. The questions focused on four aspects: a) present practices; b) attitudes, and awareness; c) priorities and d) requirements for new infrastructures, services and support. In total, participants were asked to respond to 40 questions, most of which required them to choose from a range of options and allowed a short comment, but some questions asked participants to provide their own descriptions or open narrative answers, such as to describe the nature of their digital and non-digital data sets, and explain data management and preservation issues that that face in their research.

Email invitations to complete the online survey were initially distributed via Deputy Vice-Chancellors Research inviting participation from all academic staff, research students and research support staff at ten universities in New South Wales, Australia. Further invitations and reminders were sent directly to some interest groups, centres and faculties within universities. The invitation clearly stated that "we are interested in your research practices and opinions, whatever your discipline, and whatever the extent of ICT use in your research", thus targeting both researchers who may be new to eResearch as well as those more familiar. Participation was voluntary, response rates low (760 of a possible 32,843 full and part time staff at New South Wales universities in 2012 (Australian Government Department of Education and Training, n.d.)) and, as such, the sample does not represent the whole academic population of Australian or

2 DCC: http://www.dcc.ac.uk/
3 ANDS: http://www.ands.org.au/

 doi:10.2218/ijdc.v10i2.329

New South Wales universities, but rather provides an indication of interested individuals opinions at the particular point in time.

Nevertheless, the study sought responses from a very broad university audience, and participants' answers provide a useful insight into the researchers' perspective on current research practices and challenges that are broadly associated with the development of digital research infrastructures and emergence of new research practices (Markauskaite, Hellmers, Richardson and Kennan, 2012).

This paper focuses on the responses to the multiple choice and open survey questions about data, including, data collection, analysis, sharing, management, storage and preservation. Data analysis was conducted using exploratory and summative data analysis methods. Responses to multiple choice questions were analysed using descriptive statistics and visualisation techniques (Myatt, 2007). Written answers to open questions were analysed using several thematic analysis methods (Hsieh and Shannon, 2005; Ryan and Bernard, 2003). First, open responses were explored using word frequency analysis and visualisation techniques. After screening and taking away general words that participants used to describe various kinds of data (e.g., data, datasets, measurements, records, and files), stemmed word frequencies and synonym-based word frequencies were calculated and visualized with NVivo 10 software[4]. Second, participants' comments and answers to the survey questions relevant to the subject of data handling practices were categorised into themes using thematic content analysis techniques and explored for common patterns (Ryan and Bernard, 2003).

Findings

Participants' Background

In total 760 respondents started the survey, of which 634 (83.4%) completed at least one survey section. In total, 539 (70.9%) of the respondents completed the entire survey and the remaining 221 (29.1%) answered only selected questions. Thus, numbers and percentages are reported on a question by question basis. Responses per institution varied from small (n=20) to substantial (n=118), with the median 48 responses.

A variety of disciplines were represented in the sample (Table 1). The largest percentage of respondents reported Medical and Health Sciences as one of their major disciplines (21.2%). Other disciplines with a 10% or higher representation were Education (12.9%), Social Sciences and Humanities (11.2%), and Information and Communication Technologies (ICT) (10.0%). Some participants chose "Other" category (3.8%), with Design (n=7; 1.1%) being the most frequent "Other" answer. About one quarter of respondents (n=166 of the 628 who answered this question, or 26.4%) to this question indicated more than one disciplinary category. For 18.7% (n=31) of these respondents who had multidisciplinary interests one of the disciplinary categories was ICT, which was indicative to their professional interests in technologies and, likely, eResearch.

4 NVivo 10: http://help-nv10.qsrinternational.com/desktop/procedures/run_a_word_frequency_query.htm

Overall, the respondents' familiarity with eResearch domain varied. To the question "Prior to this survey, had you heard of the term 'eResearch'?" just more than half of the respondents (52.2%) answered "Yes", while the remaining participants (47.8%) answered "No". This was despite the fact that, in the Australian academic context, 'eResearch' is a commonly used word that refers to the application of advanced information and communication technologies for enhancing existing research practices (Markauskaite, Kennan et al., 2012).

Table 1. Disciplines represented in the study sample (n=628). Note: Participants were allowed to choose multiple answers, thus the sum of individual responses is not equal to the total number of participants.

Disciplinary area	n	%
Medical and health sciences	133	21.2%
Education	81	12.9%
Social sciences, umanities and arts (general)	70	11.1%
Information, computing and communication sciences	63	10.0%
Behavioural and cognitive sciences	53	8.4%
Biological sciences	50	8.0%
Engineering and technology	50	8.0%
Commerce, management, tourism and services	41	6.5%
Law, justice and law enforcement	33	5.3%
Science (general)	30	4.8%
Studies in human society	29	4.6%
Agricultural, veterinary and environmental sciences	28	4.5%
Language and culture	26	4.1%
Chemical sciences	25	4.0%
The arts	24	3.8%
Physical sciences	23	3.7%
History and archaeology	23	3.7%
Earth sciences	22	3.5%
Economics	21	3.3%
Mathematical sciences	19	3.0%
Philosophy and religion	11	1.8%
Architecture, urban environment and building	9	1.4%
Policy and political science	9	1.4%
Journalism, librarianship and curatorial studies	6	1.0%
Other	24	3.8%
Total	628	100.0%

Of the respondents, 70.2% were academic staff, 20.6% were postgraduate research students and 6.0% were "other", the latter comprised mainly of research support staff, information technology (IT) staff, clinical staff and librarians (Figure 1). The remaining 3.2% were visiting and honorary academics. Early career academics were slightly better represented in the survey (26.0%) than mid-career (22.8%) or senior academics (21.4%).

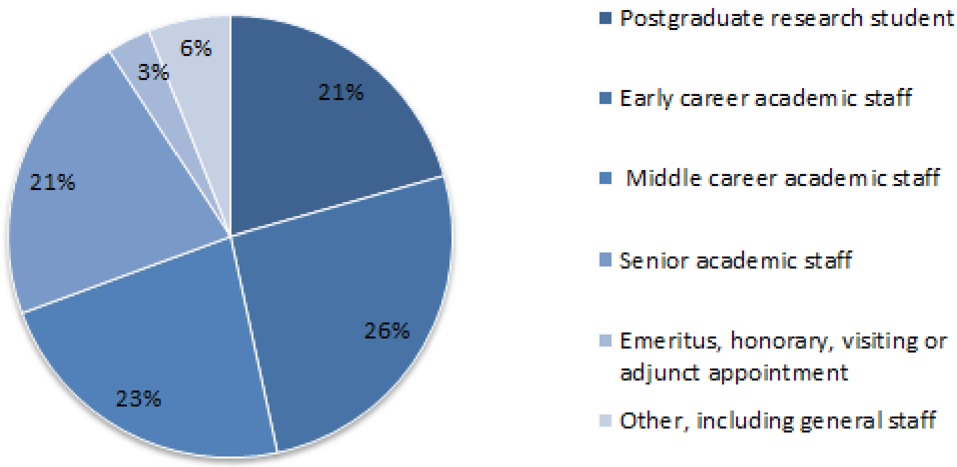

Figure 1. Participants' primary role (n=631).

Nature and Handling of Data

One aspect the study aimed to explore was what kinds of data researchers collected. The question requested respondents to list up to three words or phrases that describe the nature of their main digital and non-digital data. From the total of 493 participants who responded to this question, 73 filled in only one field, 167 two fields and 253 all three fields providing a total of 1,166 responses. Some respondents indicated general information about their data (e.g. qualitative or numerical data) and some others described the nature of their data very specifically (e.g. pottery descriptions, CAD/BIM models). After screening frequencies of stemmed words were calculated and visualized (Figure 2). In total, 164 common words were identified (a frequency of two and above). The most frequently mentioned data types were surveys (195 occurrences) and interviews (194 occurrences). This was followed by laboratory data (83 occurrences), observations (76 occurrences), photographs (48 occurrences) and behavioural data (44 occurrences). Other often mentioned data types included documents, imaging, video and sequencing data (each had 21-30 occurrences), followed by news, simulations and field data (each had 16-20 occurrences). Taken together, the above words comprised about half of keywords used by the participants to describe their data. The remaining often referred to various kinds of textual and media data, particularly data that are available online (e.g., literature, articles, linguistic data, archival records) and a range of specialized data specific disciplinary fields (e.g., modelling data, X-rays, genotyping, financial data, administrative records).

Figure 2. Main words used to describe the nature of research data.

Respondents were also asked what proportion of their research data and materials were digital (Table 2). The majority of participants (93.2%) indicated that at least 20% of their data was digital and nearly half of them (49.6%) responded that almost all their research data (81%-100%) is digital.

Table 2. Percentage of research material and data that is digital (n=557).

Proportion of digital data	n	%
0%-20%	38	6.8%
21%-40%	48	8.6%
41%-60%	67	12.0%
61%-80%	128	23.0%
81%-100%	276	49.6%
Total	557	100.0%

Respondents were also asked what kinds of software they used for their data handling and analysis (Table 3). Given the predominance of survey data and laboratory measurements it is not surprising that the most commonly used data handling and analysis tools are spreadsheets, databases and statistical software, which are used at least occasionally by at least two thirds of participants (69.2%-90.5%), and often by more than half of them (41.4%-53.4%). However, despite the high proportion of interviews, observations, photographs, videos and other qualitative data, software for analysing these kinds of digital materials has much lower usage. The most often used

 doi:10.2218/ijdc.v10i2.329

analytical tools are general qualitative analysis software and biometric software, which are used at least occasionally by about 40% of respondents, but less than half of them (12.4%-18.9%) use these software tools often. Despite the dominance of interviews and video data, the use of digital voice recognition and transcription tools, as well as special software for analysis of video and audio data was unexpectedly low, with less than one third of participants responding that they use these tools at least occasionally (21.4%-31.0%) and fewer than 10% of them responding that they do this often (6.8%-9.1%). The low use rate of digital voice recognition software or transcription tools may reflect the still developing state of the digital tools available at the time of the survey in terms of accuracy and the use of transcription services. Despite many developments, automatic speech recognition is an extremely complex problem and many developed automatic transcription tools do not perform sufficiently well for practical purposes.

It is interesting to note that a relatively large proportion of participants (39.5%) answered said that they at least occasionally used software specifically developed for the analysis of their data, and more than half of these participants (22.7%) indicated that they use this special software often. However, given the diversity of specialised data mentioned by the participants before (Figure 2), this outcome is not so surprising.

Table 3. Commonly used data handling and analysis tools (n=601).

Software used for data analysis	Use often	Use occasionally	Don't use
Spreadsheets	53.4%	37.1%	9.5%
Databases	41.9%	39.3%	18.7%
Statistical or numeric computation software	41.4%	27.8%	30.8%
General qualitative data analysis software	12.4%	28.8%	58.8%
Software specifically developed for your data	22.7%	16.8%	60.5%
Bibliometric or infometric software	18.9%	19.6%	61.6%
Visualisation software	13.6%	23.3%	63.1%
Modeling or simulation software	15.2%	21.3%	63.6%
Digital voice recognition and transcription	9.1%	21.9%	69.0%
Special visual or audio analysis software	6.8%	14.6%	78.6%
GIS or other spatial software	5.6%	11.6%	82.8%
Data mining software	5.2%	9.6%	85.2%
Other	6.6%	2.6%	90.9%

Obtaining and Sharing Data

Respondents were asked how often, and in which ways, they obtained their data. A very large majority of participants at least occasionally (95.3%) collected or created their data by themselves, and most of the participants (76.9%) did this often (Table 4). Many participants (88.2%) also answered that they at least occasionally collected data as a part of their research teams, and more than half of the participants did this often (53.6%). More than half of the respondents also obtained data from other researchers (65.4%), third party organisations (58.5%), and data archives or repositories (54.4%). However, many of these participants obtained their data in these ways only occasionally (50.9%, 41.5% and 35.3% respectively) and very few them did this on a frequent basis (14.5%, 17.0% and 19.1% respectively). A large majority of the participants (73.1%) never used a commercial data source and only about a quarter of researchers (27.1%) obtained their data from a commercial source at least occasionally. However, only 9.2% of the respondents did this often.

Table 4. Common ways of acquiring data (n=557).

Ways for acquiring data	Use often	Use occasionally	Don't use
Collected or created by yourself	76.9%	18.4%	4.7%
Collected or created as part of a team	53.6%	34.6%	11.8%
Acquired from another researcher or team	14.5%	50.9%	34.6%
Acquired from third party research organisations	17.0%	41.5%	41.5%
Acquired from academic data archives or repositories	19.1%	35.3%	45.6%
Acquired from commercial online sources	9.2%	17.9%	72.9%
Other	2.1%	2.3%	95.6%

Researchers were then asked whether, once they had collected their own data, they would be willing to share them outside of their research team or project. While more than half (54.7%) indicated that they would not be prepared to share any of their data, 36.4% indicated they would be prepared to share some of their data and 8.9% indicated that they would be prepared to share most of their data.

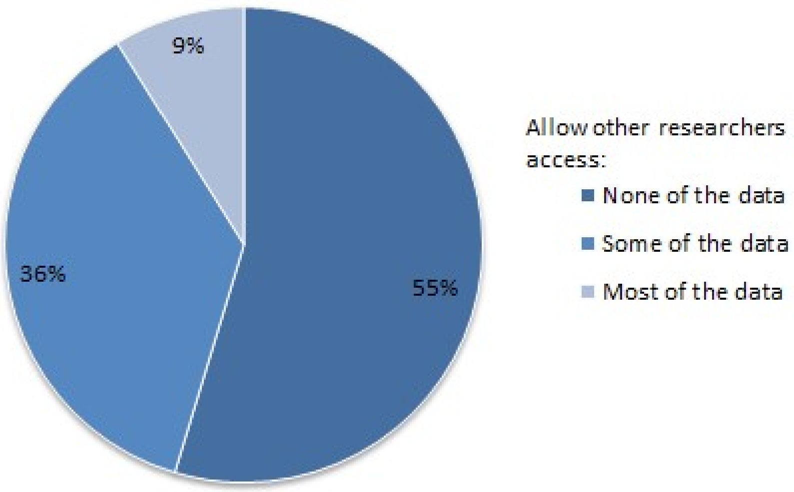

Figure 3. Researchers' readiness to share data (n=541). Note: Original question "Do you allow researchers from outside your team/project to access your research data?"

Of those who were sharing at least some of their data (n=258), the majority (72.6%) said that they enable access to their data via private negotiation (Table 5). However, less than one third of the participants responded that they would publish their data online, for example via a project website (31.9%), or submit for publishing with papers in journals (25.4%). Very few participants answered that they deposit their data in open repositories (15.7%) and even fewer said that they provide access to their data via a third party, such as a funding body, institution or experimental facility (11.3%). This is perhaps not so surprising, as at the time of the survey, while all Australian universities had institutional repositories, their main focus was on publications and the specialist data archives, such as the Australian Data Archive (re-established in 2011), were rare. They are now on the increase.

 doi:10.2218/ijdc.v10i2.329

Table 5. Ways in which researchers typically provide access to their research data to people outside their research teams (n=258). Note: Participants were allowed to choose multiple answers, thus the sum of individual responses is not equal to the total number of participants.

Ways of providing access to data	n	%
Allow access data via privately negotiated access	180	72.60%
Publish data online (e.g. on a public project website)	79	31.90%
Submit data for publishing when publish papers in eJournals	63	25.40%
Deposit data to open data repositories	39	15.70%
Access to data is provided by a third party	28	11.30%
Other	13	5.20%
Total	258	100%

Those who indicated they were not willing to share their data (n=323) had many reasons for restricting access (Table 6), but among most common reasons indicated by the respondents were issues related to privacy and confidentiality (59.1%) and ethics (29.1%). The high proportion of respondents from medical and health sciences, education and the social sciences (Table 1), along with the predominance of interview, survey and observation methods reported (Figure 2), increase the likelihood of ethical issues such as consent and data sensitivity being perceived as a barrier. Competitive research advantage and commercialisation potential were also important concerns for a large proportion of the respondents, 39.9% and 18.3% respectively. Other barriers included the technical difficulties of making data available (22.9%), lack of incentive to make data available (21.1%), and finally a perception of lack of usefulness of their data to others (14.2%). While these technical and motivational issues were mentioned by a slightly smaller number of participants, nevertheless they are important obstacles for not sharing. Interestingly, 8.7% of the respondents said that they have no reason for not sharing their data.

Table 6. Researchers' reasons for restricting access to their research data (n=323). Note: Participants were allowed to choose multiple answers, thus the sum of individual responses is not equal to the total number of participants.

Reasons for restrictions	n	%
Privacy and confidentiality issues	191	59.1%
Competitive research advantage	129	39.9%
Other ethical issues	94	29.1%
Technical difficulty of making data available	74	22.9%
Lack of incentive to make data available	68	21.1%
Commercialisation potential	59	18.3%
Lack of usefulness of my data to others	46	14.2%
Licensing issues	32	9.9%
None	28	8.7%
Other reasons	20	6.2%
Total	323	100.0%

Data Storage

To the question "How much digital research data do you currently store?", only 3.1% of respondents reported having no stored research data (Table 7). Although most respondents have some digital research data (see also Table 2), their stored datasets were generally small. Most participants (40.2%) indicated that they stored less than 10GB of research data, and another large group (38.4%) answered that they have less than 1TB stored. Less than 20% of respondents had data sets larger than 1TB. Yet, the majority of them (13.2%) had less than 10TB and only 4.9% of respondents had more than 10TB.

Table 7. Amount of digital research data currently stored by researchers (n=555).

Size of stored digital research data	n	%
None	19	3.4%
Less than 10 gigabytes	223	40.2%
Less than 1 terabyte	213	38.4%
Less than 10 terabytes	73	13.2%
More than 10 terabytes	27	4.9%
Total	555	100.0%

Although many respondents reported relatively small data sets, when asked whether they needed additional storage, 31.2% (n=174) responded in the affirmative and, interestingly, 18.5% (n=103) did not know. In the free text field associated with this question, which asked "If yes, how much do you need in total?" (Table 8), 82 respondents commented on the actual amount of space they required. Of them, about one third (31.7%, n=26) respondents required up to 1TB; and another one third (32.9%, n=27) respondents required up to 10TB. Only a small number of the respondents (9.8%, n =8) indicated that their needs exceed 10TB, with about half of them giving estimates in the range of terabytes and another half indicating petabyte values. Of the remaining respondents (25.6%, n=21), many (19.5%, n=16) simply indicated that their needs were growing and they were unsure how much storage they will need in future, with comments such as "ongoing growth!", "I am not sure but video data is increasing all the time," and "grows with time." All other respondents (6.1%, n=5) commented primarily on the importance access and quality of storage, with comments such as "need access to shared server" or "quality of storage is more important than space" or "it's about having access at the university." Researchers with large storage needs perceive that meeting

Table 8. Researchers' estimates of needed storage for research data (n=82).

Needed storage	n	%
Up to 1 terabyte	26	31.7%
Up to 10 terabytes	27	32.9%
More than 10 terabytes	8	9.8%
Unsure, but needs grow	16	19.5%
Concerns about quality	5	6.1%
Total	82	100.0%

 doi:10.2218/ijdc.v10i2.329

these needs is critical. Some state, for example, that "...data stored in the less reliable digital media from the past is now lost" or that current "storage media is often questionable and risks the loss of valuable data that have been collected over the years". One respondent notes that "raw data would be very useful to archive, but [we] have not had the capacity in the past" and another that "our research participants are child patients who have large physical files – these must be kept until they are 21 so storage is a huge issue."

When asked where they stored their data during a project, respondents nominated a variety of places. Temporary and insecure local storage options featured highly, such as internal computer hard drives (80.8%), external hard drives (66.3%), USB sticks or flash drives (61.9%) and CDs or DVDs (28.4%). More secure external storage was indicated by far smaller number of participants and included such places as local area networks (35.7%), central IT or other university affiliated computing centres (21.4%) and departmental computing centres (10.1%). Of the 8.8% (n=48) of respondents who specified "Other" locations, nine respondents reported that their data was stored in or on the cloud during a project, another 20 respondents specifically mentioned Dropbox, 11 mentioned that their data was hardcopy and several specified "lab books", "note books" or "in locked filing cabinets." Two respondents mentioned that they used their home computer as back up. Many respondents (89.0%) indicated multiple storage places during a research project, possibly as an attempt to ensure the safety and back up of the data. However, a large proportion of participants (41.8%, n=228) did not have their data stored in any other media, but local storage devices.

Table 9. Data storage locations during a project (n=546). Note: Participants were allowed to choose multiple answers, thus the sum of individual responses is not equal to the total number of participants.

Storage of data during a project	n	%
Internal computer hard drive	441	80.8%
External hard drives	362	66.3%
USB/flash drives	338	61.9%
Storage on a local area network	195	35.7%
CDs or DVDs	155	28.4%
Storage within a university affiliated computing centre	117	21.4%
Storage within a departmental computing centre	55	10.1%
Commercial data storage facility	27	4.9%
Tapes	25	4.6%
State or national data centre or repository	9	1.6%
Disciplinary data centre or repository	3	0.5%
Not sure/ don't know	2	0.4%
Other	38	7.0%
Total	546	100.0%

Data Management and Preservation

This situation of data being stored on temporary and/or insecure devices is likely to continue even after projects are completed. When asked who typically stores and cares for collected data after the end of a research, many participants reported that they themselves or another project or team member was responsible (81.2%) (Table 10). Only 8.0% stored data in an internal facility or university service, and even fewer in a state or national data centre or repository (0.5%) or a disciplinary data centre or repository (0.4%). Most disturbing was that some participants answered that they were not sure who was responsible (5.6%) or reported that nobody was responsible for the data (3.3%). At the time of the survey many institutions did not have data archives or repositories or policies or support for data management planning (Corrall, Kennan and Afzal, 2013). Since the study was conducted there have been an increasing number of national, state and institutional data policy developments and infrastructure development and discovery initiatives (Burton, Groenewegen, Love, Treloar and Wilkinson, 2012; Groenewegen and Treloar, 2013).

Table 10. Who is responsible for data after a project? (n=549).

Who stores data after the end of a project?	n	%
Me or other project/team member	446	81.2%
An internal faculty or university service	44	8.0%
A state or national data centre or repository	3	0.5%
A disciplinary data centre or repository	2	0.4%
Nobody	18	3.3%
Not sure/don't know	31	5.6%
Other	5	0.9%
Total	549	100.0%

If generally data are not stored in a repository or archive, and researchers usually take care of their data themselves, the question then arises as to whether they face any data management preservation issues and how they manage their research materials and data.

Only 41.2% of participants could categorically state that they did not have any data management or preservation issues in their research. In contrast, about one third of respondents (33.7%) explicitly reported that they do have such issues and about one quarter (24.2%) were not sure whether there are management or preservation issues with their data (Figure 4).

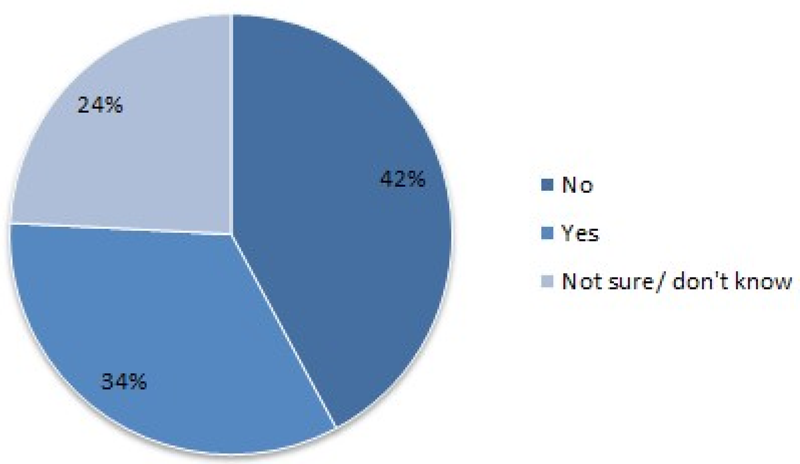

Figure 4. Do researchers face data management and preservation issues? (n=546).

One hundred and forty nine respondents described their data preservation and management issues in a free text field. The gist of these concerns mentioned by the participants is depicted Figure 5. It shows the frequencies of the stemmed words that were used by participants in their responses at least twice (192 words). The major theme that emerged from the participants' responses was related to storage, with words "storage" (52 occurrences) and "store" (26 occurrences) mentioned most often, and a number of other frequently mentioned words related to this topic, such as "backup" (15 occurrences) and "space" (14 occurrences). This was followed by words referring to "access" (26 occurrences), and group of words referring to data management and preservation, such as "filing" (25 occurrences), "archiving" (23 occurrences) and "management" (22 occurrences). Researchers referred to different time periods and scales (e.g., "current" "now", "back" and "long"); and various institutional structures (e.g., "university," "centralized" and "faculty").

Figure 5. Main words used to describe data management and preservation issues.

In order to get more accurate insight into the challenges faced by researchers we conducted thematic analysis of the participants' responses. The major issues raised by respondents fall into six categories:

- Hosting or storage;
- Maintenance of integrity of the data;
- Access and retrieval;
- Security;
- Ethical and intellectual property issues; and finally,
- Human resources.

The major category that emerged, mentioned explicitly by 85 respondents, was the need for assistance with data storage. This was variously expressed by respondents as a need for "logical and meaningful storage", or a "need [for] data to be organizationally retained and archived" or the "need [for] data to be stored centrally." Often mentioned with this need for managed storage or hosting was the issue of maintaining the integrity of digital data (n=25). Using terms such as "technical obsolescence"; "format rot"; "fragile" data; "short term shelf life" of the hardware on which they stored their data; "incompatibility" of old data with new software and hardware, the respondents indicated that they need help in this domain, with eight of these reporting an actual accidental data loss. The third major issue (n=24) arising was managing access to, and later retrieval of, the data. Several respondents talked about "rubbish" or "missing metadata" or their lack of knowledge about knowledge representation and file structures sometimes making re-finding their own data difficult. Some participants specifically mentioned issues about the difficulties of finding data from past students and staff, with one plaintive appeal "what happens after I retire?" In these latter responses the focus on access appears to relate to access and future use by researchers of their own or their colleagues older data, rather than opening up access to data to external researchers or the public. These responses appear to indicate that researchers' priorities for their data still seem to centre on their own use of their data rather than encompassing wider concerns about the public sharing of data.

Some respondents (n=24) mentioned their concerns with the security of their data. They expressed these concerns in terms of worries about safety of storage on the cloud and the common practice of sending data via emails, and others in terms of "data corruption" and one even in terms of "sabotage." While many participants mentioned privacy, confidentiality and ethical issues as well as concerns about the competitive research advantage as the main obstacles for sharing their research data, only ten respondents reported their concerns in managing ethical and intellectual property requirements. This surprising result suggests that many researchers perhaps do not think that these obstacles for data sharing could be solved by putting in place better data management and preservation arrangements. The final issue raised was the need for human resources – people to assist in this regard (n=12). Respondents talked about the lack of expertise in managing and preserving their data, one saying "I'm a shambles" and another "I'm disorganized. There are only 24 hours in a day."

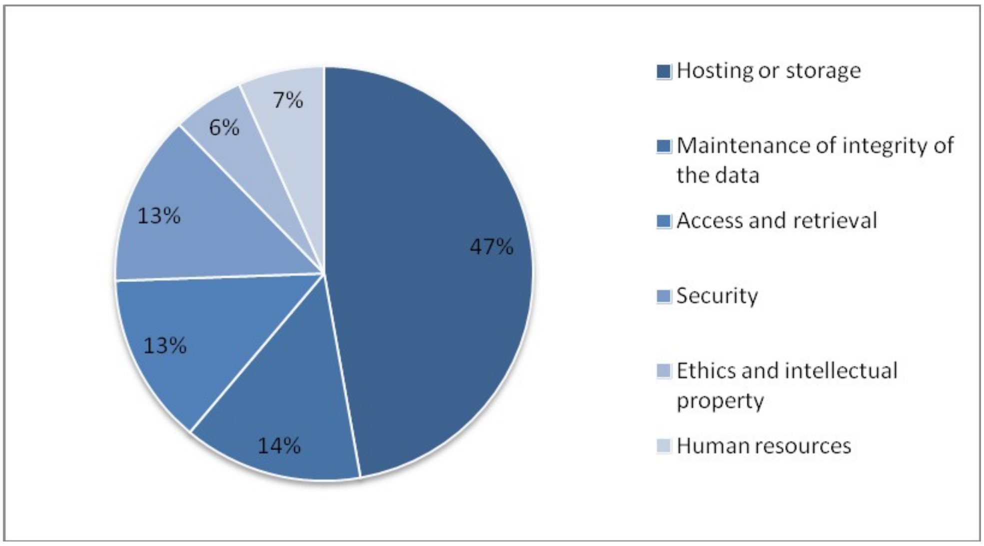

Figure 6. Data management and preservation issues identified by researchers. Note: Some participants indicated several issues, thus the sum of individual responses is not equal to the total number of participants (n=145) who answered this question.

Despite so many researchers being responsible for their own data, and identifying that they need assistance with managing it, when asked whether they have an explicit data management plan, only about one quarter (27.5%) were able to respond that they did. A majority (57.5%) did not have an explicit data management plan, and the rest (15.1%) were unsure whether projects they were involved with had data management plans.

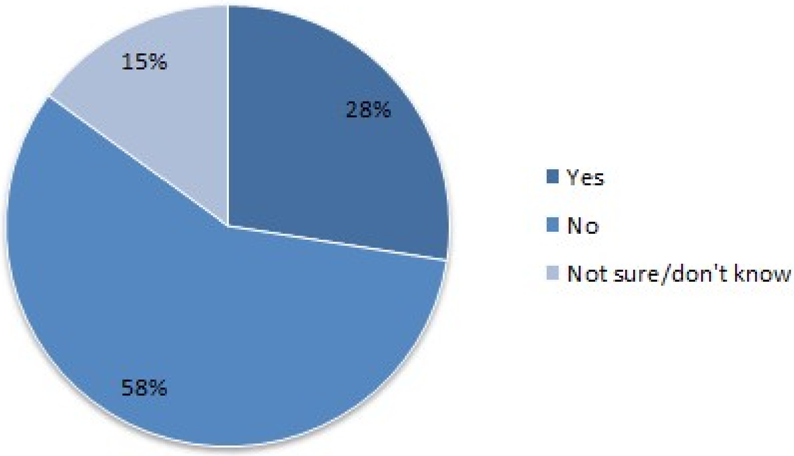

Figure 7. Do researchers have an explicit data management plan? (n=546).

As discussed earlier, at the time this survey was conducted, few Australian institutions nor the major Australian research funding bodies had research data management policies and not all policies current then, or developed since, require researchers to produce formal research data management plans. However, data management plans are becoming more common as funders, institutions and researchers themselves see data management planning as a way of providing a useful and systematic process for documenting the activities required to improve access to, and longevity, sharing and reuse of research data, while also managing and anticipating risk and clarifying ethical and intellectual property issues (Donnelly, 2012).

Discussion and Conclusions

In this paper we shed light on practices of researchers at the time of the survey. From our sample of university-based researchers we have described a wide breadth of disciplines working with data and the extremely diverse nature of data collected, created, used and requiring management in the university setting, reflecting findings reported earlier in the literature (Borgman, 2007; Charlton et al., 2013; Scifleet, Henninger and Albright, 2013). While the majority of data were digital, there were still substantial pockets of analogue data which must also be considered in institutional and funder policies and research data management plans. The numbers of researchers requiring large amounts of space were small (80% of our respondents could probably fit their data on the average external hard drive), but those with large data storage needs perceived these needs as critical and report them as not being met. It would be interesting for future research to test the hypothesis that smaller data users are better able to manage their own storage needs locally, whereas bigger data users rely on institutional or departmental networked storage? The alternative hypothesis might be that bigger data users have had to develop their own local storage solutions in order to manage their more complex storage requirements? Understanding the answers to these questions would enable institutions and other data archive developers to understand whether they should focus on meeting the relatively small storage needs of the many, or focus on supporting the smaller number of researchers who have larger storage requirements, where the bigger data might have been more costly to create and the impact of their loss due to insecure storage be greater?

In addition to there being a broad range of data types, there are broad range of softwares used to handle and analyse data. It is also noted that a surprising proportion (nearly 40%) of respondents occasionally used software specifically created for their data or a project, for analysis. This may create issues for storing and sharing, particularly if software, and particularly specifically created software, is required for future re-use and analysis.

While most respondents had only relative small data sets, many reported their during-project data management as temporary and insecure, with examples of email exchanges and data on temporary devices, such as laptops, USBs and external drives, as problematical. After project storage, archiving and maintenance of data integrity are also recognised as issues, though fewer researchers actually mentioned post-use lifecycle issues. While it is recognised that large data sets pose particular issues for researchers, it is recognised by many respondents that even smaller data collections require a storage “home” for data, which if not a space where the data is formally archived and curated or open and able to be shared, is at least a space where researchers can be confident that their data will remain for the useful life, or ethical requirement, of the data. It was evident from the survey responses that many researchers felt such storage, let alone curation services, were not available to them despite the increasing rate of institutional policies and procedures for data management by their institutions (Corrall et al., 2013; Groenewegen and Treloar, 2013).

While the vast majority of respondents create or collect their own research data there are a wide variety of other reported ways of obtaining data. While just more than half of the respondents had occasionally acquired data from other researchers, 41% from third party organisations and 35.3% from data archives, only 36.4% were willing to share their own data and reported a wide range of reasons for not sharing. As researchers predominantly collect or create their own data and only occasionally use curated data, researchers are unlikely to see the value in spending time managing and curating their own data or depositing it in a data repository for sharing, or even preservation and archiving. The reasons for not sharing may change as the potential of aggregated data or reusing data become increasingly recognised, and as institutions and funders increasingly require that data management plans be explicit and data be findable. Also, as researchers are more likely to share informally they possibly have the ability to interact directly with the data collector/creator to ask questions about provenance and methodologies that might be provided via metadata in an archive. Thus maintaining a good network of contacts in their field may be considered a more important and higher priority activity than managing their data in order to share with unknown reusers in unknown contexts and this is possibly reflected in our finding that many plan to share their data via negotiated access. It would be interesting in future studies to explore this issue in more detail to discover whether researchers who typically reuse existing data more frequently would typically share their data by a similar method to how they obtained their own data e.g. personal contact, online via project websites, via deposit in data repositories. However, regardless of whether data are to be shared, safe storage was reported as a continuing issue.

In addition to the provision of infrastructure, such as storage, and services, such as curation for large nationally significant datasets funded by the government, there are also issues for researchers about which further discussion and/or education is necessary. These include issues identified by researchers in the survey as reasons for not sharing data, such as how to manage ethical concerns, privacy and confidentiality; how to make decisions about the future usefulness of data; and how to provide support for smaller data sets with technical issues, such as metadata creation for discoverability (future access) and curation and preservation across the lifecycle to the prevent obsolescence and loss. Data management planning at the outset of a project may assist researchers in identifying and systematically addressing these issues at the early stages of a project.

Despite many respondents reporting issues with their data, only 27.5% reported that they had an explicit data management plan. While incentives to plan to manage data more rigorously may change due to the raising awareness of the potential uses of shared, aggregated and reusable data through advocacy work conducted by organisations such as the Australian National Data Service (ANDS), the State based eResearch agencies, international agencies such as DCC and DataOne and with the proliferation of institutional and funder policies about data management, many of the data management issues reported by researchers in this survey will not be resolved by policies. We therefore recommend that in addition to policy work that wide discussion is undertaken, and where necessary supplemented by education and training, about how more formalised research data management planning can help researchers resolve some of the issues raised by them in this survey. Ethical issues can be mitigated as planning can indicate early on that, for example, gaining appropriate consent prior to the conduct of the research is essential for sharing, or by planning to anonymise to retain less sensitive subsets of data. Our own experience reflects this. For this survey we did not explicitly ask for consent via the ethics approval process to publicly share the survey data, thus under the conditions of that consent we are only able to make an anonymised aggregated summary available (Markauskaite, Hellmers et al., 2012). Future surveys conducted by one of the authors have learned from this scenario, the appropriate consents have been requested and data shared (Kennan, Corrall and Afzal, 2015).

A plan can also assist researchers to estimate in advance how much and what kinds of storage they are likely to need at various stages of their research and to address early on issues identified about future access. Planning encourages researchers to think about their data throughout potentially useful lifecycle of the data. The information in research project or programme data management plans could also possibly be captured and used by institutions to inform their plans to develop institutional storage infrastructures and to identify skills gaps, enabling institutions to, for example, recruit data specialists or plan to undertake training as required.

It was noted by participants that infrastructure such as storage need to be provided. In the 2009/10 budget the Australian Government funded a national network of research data storage and collaboration infrastructure to be built by the Australian Research Collaboration Service ($97 million) (Australian Government, 2010). It will be interesting to note in future surveys how the development of this network changes researcher perceptions about the availability and suitability of research data storage. Since the study was conducted there have been an increasing number of national state and institutional data infrastructure and discovery initiatives. Institutions responsible for nationally significant collections have been supported in making those collections connected, visible and available, often through Research Data Storage Infrastructure (RDSI) nodes (Burton, Groenewegen, Love, Treloar and Wilkinson, 2012; Groenewegen and Treloar, 2013) and ANDS' Research Data Australia facility. Future studies are planned and it will be interesting to see how these initiatives are beginning to address researcher concerns.

Acknowledgement

The work reported in this paper was initiated and part-funded by University of Sydney ICT (Tools and Frameworks for Research Collaboration project) and Intersect Australia Ltd. Collaborators in the survey development included Leonie Hellmers (Intersect) and Jim Richardson (University of Sydney). We would like to thank DVCs/PVCs for Research of all participating universities for their cooperation, and all participants in this survey for their time and willingness to contribute. We also acknowledge the IJDC reviewers. Any errors and opinions are all our own.

References

Andersson, E., & Sørvik, G. O. (2013). Reality lost? Re-use of qualitative data in classroom video studies. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research,14*(3). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs130313

Australian Government. (2010). *Budget 2009-10 expense measure: Innovation, industry, science and research.* Retrieved from Australian Government website: http://www.budget.gov.au/2009-10/content/bp2/html/bp2_expense-20.htm

Australian Government. (n.d.). *Selected higher education statistics – 2012 staff data.* Retrieved from Australian Government Department of Education and Training website: http://www.education.gov.au/selected-higher-education-statistics-2012-staff-data

Australian Research Council. (2013). *Discovery projects: Instructions to applicants for funding commencing in 2014*. Canberra: Australian Government.

Australian Research Council. (2014a). *Discovery projects: Instructions to applicants for funding commencing in 2015.* Canberra: Australian Government.

Australian Research Council. (2014b). *Funding rules for schemes under the Discovery Program for the years 2014 and 2015.* Retrieved from the ARC website: http://archive.arc.gov.au/archive_files/Funded%20Research/1%20Discovery%20Programme/Discovery%20Projects/2015/DP15_Funding_Rules.pdf

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J. ... Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems, 29*(2), 599–611. doi:10.1016/j.future.2011.08.004

Beitz, A., Groenewegen, D., Harboe-Ree, C., Macmillan, W., & Searle, S. (2013). Case study 3: Monash University, a strategic approach. In G. Pryor, S. Jones, & A. Whyte (Eds.), *Delivering Research Data Management Services* (pp. 163–189). London: Facet.

Borgman, C. L. (2006). What can studies of e-learning teach us about collaboration in e-research? Some findings from digital library studies. *Computer Supported Cooperative Work: The Journal of Collaborative Computing, 15*(4), 359–383. doi:10.1007/s10606-006-9024-1

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Cambridge: The MIT Press.

Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*(6), 1059–1078. doi:10.1002/asi.22634

Burton, A., Groenewegen, D., Love, C., Treloar, A., & Wilkinson, R. (2012). Making research data available in Australia. *Intelligent Systems, IEEE, 27*(3), 40–43. doi:10.1109/MIS.2012.57

Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an approach for data management education: A report from the Data Information Literacy Project. *International Journal of Digital Curation, 8*(1), 204–217. doi:10.2218/ijdc.v8i1.254

Charlton, P., Mavrikis, M., & Katsifli, D. (2013). The potential of learning analytics and big data. *Ariadne, 71*. Retrieved from http://www.ariadne.ac.uk/issue71/charlton-et-al

Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and research data management services: Emerging trends in library support for research. *Library Trends, 61*(3), 636–674. doi:10.1353/lib.2013.0005

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368*(1926), 4023–4038. doi:10.1098/rsta.2010.0165

Data Documentation Intitiative (DDI). (n.d.). What is DDI. Retrieved from DDI website: http://www.ddialliance.org/what

Digital Curation Centre (DCC). (n.d.). DCC curation lifecycle model. Retrieved from DCC website: http://www.dcc.ac.uk/resources/curation-lifecycle-model

Donnelly, M. (2012). Data management plans and planning. In G. Pryor (Ed.), *Managing research data* (pp. 83–103). London: Facet.

Glavic, B. (2014). Big data provenance: Challenges and implications for benchmarking. In T. Rabl, M. Poess, C. Baru, & H.-A. Jacobsen (Eds.), *Specifying Big Data Benchmarks, 8163*, 72–80. Berlin: Springer. doi:10.1007/978-3-642-53974-9_7

Godbold, N. (2013). Listening to bodies and watching machines: Developing health information skills, tools and services for people living with chronic kidney disease. *Australian Academic & Research Libraries, 44*(1), 14-28. doi:10.1080/00048623.2013.773859

Grey, J. (2009). Jim Grey on e-science: A transformed scientific method. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The fourth paradigm: Data-intensive scientific discovery* (pp. xix–xxxiii). Redmond, WA: Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf

Groenewegen, D., & Treloar, A. (2013). Adding value by taking a national and institutional approach to research data: The ANDS experience. *International Journal of Digital Curation, 8*(2), 89–98. doi:10.2218/ijdc.v8i2.274

Harvey, R. (2010). *Digital curation: A how to do it manual*. New York: Neal-Schuman Publishers.

Henty, M., Weaver, B., Bradbury, S. J., & Porter, S. (2008). *Investigating data management practices in Australian universities.* Canberra: Australian Partnership for Sustainable Repositories.

Hey, T., Tansley, S., & Tolle, K. (2009). *The fourth paradigm: Data-intensive scientific discovery.* Redmond, WA: Microsoft Research.

Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation, 3*(1), 134–140. doi:10.2218/ijdc.v3i1.48

Higgins, S. (2012). The lifecycle of data management. In G. Pryor (Ed.), *Managing research data* (pp. 17–45). London: Facet.

Houghton, J. W., Steele, C., & Henty, M. (2004). Research practices and scholarly communication in the digital environment. *Learned publishing, 17*(3), 231–249. doi:10.1087/095315104323159667

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. doi:10.1177/1049732305276687

Kennan, M. A., Corrall, S., Afzal, W. (2015). *Academic library survey responses data: Research support services, including bibliometrics and research data management* [Data set]. Charles Sturt University and University of Pittsburgh. doi:10.4225/43/54ea72fdcec36

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37. doi:10.3102/0013189X12459803

Markauskaite, L. (2010). Digital media, technologies and scholarship: Some shapes of eResearch in educational inquiry. *The Australian Educational Researcher, 37*(4), 79–101. doi:10.1007/BF03216938

Markauskaite, L., Hellmers, L., Kennan, M. A., & Richardson, J. (2009). *eResearch practices, barriers and needs for support: Preliminary study findings from four NSW universities*. Paper presented at the 3rd eResearch Australasia Conference, Sydney. Retrieved from http://www.eresearch.edu.au/docs/2009/era09_submission_55.pdf

Markauskaite, L., Kennan, M. A., Richardson, J., Aditomo, A., & Hellmers, L. (2012). Investigating eResearch: collaboration practices and future challenges. In A. Juan, T. Daradoumis, M. Roca, S. Grasman, & J. Faulin (Eds.), *Collaborative and distributed e-Research: Innovations in technologies, strategies and applications* (pp. 1–33). Hershey, PA: IGI. doi:10.4018/978-1-4666-0125-3.ch001

Markauskaite, L. Hellmers, L., Richardson, J., & Kennan, M.A. (2012). *Survey: eResearch infrastructure, practices and requirements* [Data set summary]. Sydney: Intersect Australia. Retrieved from http://www.intersect.org.au/docs/eResearch_requirments_2011_responses%20responses.pdf

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: Supporting ecology as a data-intensive science. *Trends in Ecology & Evolution, 27*(2), 85–93. doi:10.1016/j.tree.2011.11.016

Murray-Rust, P. (2008). Open data in science. *Serials Review, 34*(1), 52–64. doi:10.1080/00987913.2008.10765152

Myatt, G. J. (2007). *Making sense of data: A practical guide to exploratory data analysis and data mining*. Hoboken, NJ: Wiley. doi:10.1002/0470101024

NHMRC. (2014). *Principles for accessing and using publicly-funded data for health research.* Canberra: Australian Government. Retrieved from https://consultations.nhmrc.gov.au/files/consultations/drafts/draftprinciplesaccessingpubliclyfundeddata141209.pdf

NHMRC, ARC, & AVCC. (2007, updated 2014, March). *National statement on ethical conduct in human research.* Canberra: Australian Government. Retrieved from NHMRC website: https://www.nhmrc.gov.au/guidelines-publications/e72

NHMRC, ARC, & Universities Australia. (2007). *Australian code for the responsible conduct of research*. Retrieved from NHMRC website: http://www.nhmrc.gov.au/guidelines/publications/r39

Pampel, H., & Dallmeier-Tiessen, S. (2014). Open research data: From vision to practice. In S. Bartling & S. Friesike (Eds.), *Opening Science* (pp. 213–224). Berlin: Springer. doi:10.1007/978-3-319-00026-8_14

Rice, R. (2009). *DISC-UK DataShare project: Final report.* Retrieved from Jisc repository: http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf

Pryor, G. (Ed.). (2012). *Managing research data*. London: Facet.

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods, 15*(1), 85–109. doi:10.1177/1525822x02239569

Scifleet, P., Henninger, M., & Albright, K. H. (2013). When social media are your source. *Information research, 18*(3), C41. Retrieved from http://www.informationr.net/ir/18-3/colis/paperC41.html

Schneiderman, R., Deng, L., & Sejnoha, V. (2015). Accuracy, apps advance speech recognition. *IEEE Signal Processing Magazine, 32*(1), 12–15. doi:10.1109/MSP.2014.2358717

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS ONE, 6*(6), e21101. doi:10.1371/journal.pone.0021101

Treloar, A., Choudhury, G., & Michener, W. (2012). Contrasting national research data strategies: Australia and the USA. In G. Pryor (Ed.), *Managing Research Data* (pp. 173–203). London: Facet.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society, 15*(3), 133–148.

Wallis, J. C., Borgman, C. L., Mayernik, M. S., & Pepe, A. (2008). Moving archival practices upstream: An exploration of the life cycle of ecological sensing data in collaborative field research. *International Journal of Digital Curation, 3*(1), 114–126. doi:10.2218/ijdc.v3i1.46

Witt, M. (2009). Institutional repositories and research data curation in a distributed environment. *Library Trends, 57*(2), 191–201. doi:10.1353/lib.0.0029