Researcher data practices at Swinburne: Results of a survey

Dana McKay

22 November 2007
**Executive summary**

This report discusses the results of a survey of Swinburne researchers about their practices surrounding their research data. Areas of investigation included how they store and manage their data, whether they share data or use others’ data, and how they feel about the new ARC rules governing research data. This work was undertaken to understand the differences between larger and smaller universities’ researchers’ data practices, but also to understand what is happening “on the ground” at Swinburne with a view to better supporting data management.

**Managing research data**

Researchers are by-and-large comfortable managing their research data, however they would be grateful for secure physical and digital spaces where they could confidently keep data. Researchers are very aware of their ethical obligations where such apply, and very protective of their subjects, and their data management practices reflect this. They are also aware of their obligations to keep their research data, and see personal value in doing so; a large number of researchers have re-analysed their own data. Data is created in digital and hard-copy form, though some (but not all) data is eventually digitised.

**Sharing research data**

Researchers share data (both in terms of sharing their own and using others’) relatively frequently. When data is shared, it is most often shared among colleagues; it is unusual to put data in a repository of any kind or give it to unknown researchers. Data is shared in a digital form more often than hard copy; this suggests that data that doesn’t get digitised is simply not shared.

In general, researchers (even those who have not been involved in data sharing in the past) are not averse to data sharing in the future.

**The ARC proposal**

While researchers are not averse to sharing on their own terms, they are quite concerned about mandated sharing. Commonly voiced concerns revolve around data ownership and control and ethical confidentiality obligations. Researchers also voiced a certain amount of concern about the time and bureaucracy involved in compliance with such regulations (though this is not borne out by their previous experience with regulations). There is, however, cautious optimism about the benefits the rules could have to research culture, encouraging sharing and collaboration.
Table of contents
1. Introduction......................................................................................................................................4
2. Related work ....................................................................................................................................4
3. Methodology ....................................................................................................................................5
4. Results..............................................................................................................................................5
  4.1 Demographics ............................................................................................................................5
  4.2 Characteristics of data normally used by participants ...............................................................6
  4.3 Data Management ......................................................................................................................8
  4.5 Using data generated by others ................................................................................................13
  4.6 Sharing their own data .............................................................................................................14
  4.7 The New ARC rules .................................................................................................................17
5. Discussion and conclusions ...........................................................................................................18
  5.1 Existing data practices .............................................................................................................19
  4.2 Concerns about sharing data ....................................................................................................19
  5.3 Perceived advantages in data sharing .......................................................................................19
  5.4 Assistance Swinburne could provide now ...............................................................................20
  5.5 Conclusions ..............................................................................................................................20
6. Acknowledgements ........................................................................................................................20
7. References ......................................................................................................................................21
Appendix A: The survey ..................................................................................................................22
1. Introduction
Recently international attention has been focused on preservation of research data for re-use, particularly research that is publicly funded. Using existing research repositories is one proposed solution to the data preservation question, but it faces problems when confidential data (such as medical records or participant-identifying interviews) are part of the data.

While the need to preserve data has become entrenched in funding models, and various methods of data preservation have been bandied about as potential solutions, little work has focused on understanding the existing practices and concerns of researchers with regard to the data they use. Without understanding these concerns and practices, it will be difficult to implement an effective means of data preservation, and (if institutional repository contributions are anything to go by [1][2]) even more difficult to get researcher buy-in [3].

2. Related work
In the past few years there has been a flurry of interest in data preservation and archiving. As early as 2000, Corti [4] was expounding on the need to share qualitative research data (particularly as it is difficult to gather), and discussing the technological and ethical barriers to effective data sharing (and ways to address the ethical barriers in particular [5]).

Lyon described a model research data flow (including archiving) in 2003, recognising that researchers are reluctant to deposit data in a repository or archive due to intellectual property and confidentiality reasons, but that data sharing is not only valuable but possible [6]. Heery et al. describe the implementation of this model using e-prints institutional repository software to house a collection of crystallography data [7].

Humphrey expressed concern about how an individualistic culture may negatively impact data sharing in 2005 [8], underscoring the intellectual property concerns researchers raised with both Lyons and Corti. Humphrey and Jacobs had earlier proposed an alternative property model for data [9] to circumvent the barriers to access that intellectual property models typically throw up. Arzberger et al. recommend not a different property model, but strong regulation by funding bodies and indeed the government to ensure that as far as possible, publicly funded research data are available to the public without restriction [10].

More recently, there has been considerable work in Australia on facilitating e-research repositories. The OAK Law project investigated the legal implications of data sharing in Australia, and suggested some best practice guidelines for dealing with both intellectual property and confidentiality [11]. The APSR project included an investigation of the data sharing practices of specific groups that already engage in e-research [12], with a view to identifying key elements of e-research, and areas where institutions can help e-researchers. Given that these researchers already engage in e-research, there was little discussion of the apprehensions of newcomers to e-research surrounding intellectual property and confidentiality. The DART project investigated data practices of three pre-identified groups of researchers, and engaged more with their concerns, noting that researchers are resistant to anything that does not contribute (as they see it) to their core research and takes up their time [13].

Swinburne is a smaller institution than any of the institutions studied in the literature, and thus has smaller research groups. University size is not the only difference however—our study method varies significantly from those in other reports. Rather than investigate small groups of researchers intensively, or focus on those already practising e-research, we aimed for a broader picture of researcher data management practices, and opened our survey to all Swinburne researchers.
Responses came from researchers in a number of different disciplines and career stages, including research students, corporate researchers, and professors.

3. Methodology

To reach the largest possible number of researchers (given Swinburne’s geographical spread) an online survey methodology was chosen (this is also consistent with other research done in this area). However, given the expected diversity of researcher attitudes and practices, this survey was not limited to quantitative questions, but included a large number of free text boxes as well in an attempt to elicit the nuances of behaviour and attitudes that are missed in a 5-point Likert scale. Researchers were also asked whether they would be happy to be contacted with further questions, so that any questions that arose as a result of the survey might be further explored at a later stage.

The survey included 71 questions, however, not all researchers were presented with all questions; some questions were dependent on answers to earlier questions. The pathways through the survey and the survey questions are diagrammed in Appendix A. Unfortunately, due to some technical problems with the survey software, some questions were not displayed to any researchers.

To recruit respondents, the survey was advertised in Swinburne’s Research Bulletin, a newsletter for Swinburne research staff and graduate students, and then an official email was sent out to researchers one week before the survey closed as a reminder. To further assist in recruitment, a 30GB iPod video prize was offered in a random prize draw after the survey closed. To be eligible to win the iPod, survey respondents had to complete the survey and list an email address at which they could be contacted if they won (and give us permission to use the email address supplied for this purpose). The iPod also helps ameliorate the self-selection bias with voluntary survey responses; some respondents will have completed the survey merely to win the iPod, rather than self-selecting because they had strong feelings on the topic.

4. Results

This section will detail the results of the survey on a section-by-section basis, but conclusions will be drawn Section 4 (Discussion) of this document.

4.1 Demographics

85 respondents completed the survey. This section gives some background information about those people.

Research roles and experience

The research roles of survey respondents were as follows:

<table>
<thead>
<tr>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masters student</td>
<td>5</td>
</tr>
<tr>
<td>Doctoral Student</td>
<td>26</td>
</tr>
<tr>
<td>Postdoctoral Fellow</td>
<td>4</td>
</tr>
<tr>
<td>Academic</td>
<td>33</td>
</tr>
<tr>
<td>Consultant</td>
<td>1</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1: Research Roles

In the ‘Other’ section, four respondents gave free-text descriptions of their roles, which included management and project work. Respondents were also given the option ‘Medical Researcher’, but as none of them chose this, it is not represented here.

1 Listed in the survey as Research Fellow/Assistant
2 Clarified in the survey as lecturer/senior lecturer/professor
Two respondents were answering on behalf of their research groups as well as themselves. These groups were relatively small (both fewer than five people) and had each been running for a timeframe longer than 12 months and shorter than 10 years.

The length of time in research roles for the remaining 83 respondents is as follows:

<table>
<thead>
<tr>
<th>Time Frame</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-2 years</td>
<td>14</td>
<td>16.9</td>
</tr>
<tr>
<td>2-5 years</td>
<td>31</td>
<td>37.3</td>
</tr>
<tr>
<td>5-10 years</td>
<td>18</td>
<td>21.7</td>
</tr>
<tr>
<td>10-20 years</td>
<td>10</td>
<td>12.0</td>
</tr>
<tr>
<td>20+ years</td>
<td>10</td>
<td>12.0</td>
</tr>
</tbody>
</table>

Table 2: Length of research experience

Surprisingly, there was not a statistically significant correlation between a researcher’s role and the length of their career. However, given that at least one academic indicated their research career had been shorter than 3 years, and as Swinburne lecturers are required to have a PhD, we can see that ‘length of career’ is a somewhat subjective measure anyway.

**Fields of Study**

<table>
<thead>
<tr>
<th>Field</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>19</td>
<td>22.6</td>
</tr>
<tr>
<td>Engineering</td>
<td>8</td>
<td>9.5</td>
</tr>
<tr>
<td>IT</td>
<td>8</td>
<td>9.5</td>
</tr>
<tr>
<td>Design</td>
<td>4</td>
<td>4.8</td>
</tr>
<tr>
<td>Social science</td>
<td>27</td>
<td>32.1</td>
</tr>
<tr>
<td>Law and business</td>
<td>11</td>
<td>13.1</td>
</tr>
<tr>
<td>Education</td>
<td>7</td>
<td>8.3</td>
</tr>
</tbody>
</table>

Table 3: Research fields

Researchers were invited to enter a field of study in a free text field thus allowing them to be as vague or specific as they chose—one researcher chose not to answer this question. From the entries given a generative taxonomy was created, and each entry tagged with one class. Both the taxonomy and classifications were verified by a second party who had not seen the original taxonomy or classifications.

Science and social science fields were further divided to understand the distribution of respondents in these fields; it was found that 10 social scientists (11.9% of the total number who gave a field, and 37.0% of social scientists) declared psychology as their research field (the remainder of those who had declared a social science research interest declared broadly variant interests). 47.4% of entries classified as science (9 entries, or 10.7% of the total number of entries) were physics and astronomy related, and a further 36.8% (7 entries or 8.3% of the total number of entries) were biological and medical science related.

**4.2 Characteristics of data normally used by participants**

Survey respondents were asked about the type of data they ‘typically’ used in an attempt to gain an insight into what their data practices might be like, but also with a view to determining what kinds of data any proposed institutional level software would have to be able to cope with.

**Data types**

84 respondents answered a question about what type of data they normally used in their research, however most respondents listed more than one type of data (range 1-5 types, median 2 types, mean 2.3 types). Based on the types of data listed, several broad categories of data were created, and the distribution is shown below:
Table 4: Types of data used by researchers

| Data types classified as ‘other’ include video materials, theory, statistics and case studies. |
|---|---|---|
| **Data formats** | All 85 respondents described the formats that their data comes in normally. Again, there are more than 85 responses in total, because many researchers selected more than one data format (range 1-9 formats, median 4 formats, mean 3.6 formats). | ‘Other’ data formats include Matlab and Powerpoint documents. Most respondents used a combination of digital formats and their traditional counterparts; there was no clear split between early adopters and their more cautious counterparts. We can see here that researchers tend to use a mix of digital and non-digital data, and a significant proportion of the data described (38.6%) is in non-digital format. |
| **Other data characteristics** | Respondents were asked to name anything else that made their data unusual or different; examples given included ethics constraints and cultural issues. 26 researchers claimed some special characteristics of their data (30.6% of respondents). Within those 26 responses, 15 respondents mentioned that their data was ‘sensitive’ or ‘confidential’ or subject to ethics requirements (conversely one respondent said that his or her data was never subject to ethics or privacy concerns). Two participants mentioned that they work with an unusually large volume of data, and a further three mentioned special processing as a requirement of their data. Two participants noted they use data in a language other than English. |
4.3 Data Management

Researchers were asked a series of questions about how they manage their data, specifically with reference to the last publication they wrote.

**Data type and formatting**

Participants were asked to describe the data they used in their most recent publication; responses to this question largely reflect the type of data researchers “generally use” as described in section 3.3. Respondents were also asked specifically whether their data was all digital, partly digital or not digital at all. 50 researchers (58.8%) created only digital data, 14 (16.5%) created no digital data, and 21 researchers (24.7%) created a combination of the two.

Researchers were also asked whether their data required any computerisation or format changes before they could use it; 44 researchers (51.8%) did have to change the format of their data, 31 researchers (36.5%) did not have to change the format of their data, and the final 9 (10.6%) reformatted some of their data, with the fractions listed being between 20 and 50 percent.

<table>
<thead>
<tr>
<th>Created only digital data</th>
<th>Created no digital data</th>
<th>Created a combination of digital and non-digital data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage who</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reformatted all data</td>
<td>20.0</td>
<td>78.6</td>
</tr>
<tr>
<td>Percentage who</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reformatted no data</td>
<td>54.0</td>
<td>21.4</td>
</tr>
<tr>
<td>Percentage who</td>
<td></td>
<td></td>
</tr>
<tr>
<td>reformatted some data</td>
<td>10.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 6: Original data format and reformatting

Table 6 compares the amount of reformatting necessary between those who created all data digitally and those who created some or none digitally. As we might expect, those who create their data digitally are less likely to need to reformat the data later.

This reformatting (which is quite prevalent) can take considerable time; Table 7 describes the responses of researchers to a question about this.

<table>
<thead>
<tr>
<th>Very unsure</th>
<th>&lt;1 hour</th>
<th>1-2 hours</th>
<th>~1 day</th>
<th>~1 week</th>
<th>~1 month</th>
<th>&gt;1 month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>12</td>
<td>0</td>
<td>5</td>
<td>9</td>
<td>17</td>
<td>5</td>
</tr>
<tr>
<td>Percentage</td>
<td>22.2</td>
<td>0.0</td>
<td>9.3</td>
<td>16.7</td>
<td>31.5</td>
<td>9.3</td>
</tr>
</tbody>
</table>

Table 7: Time taken to reformat data

Researchers were asked to indicate how long data processing took, if it was more than one month. Two researchers mentioned ongoing work for a certain amount of time each week; in all other cases the range was 2-4 months.

Researchers who had to work with their data to make it suitable for analysis were asked what that work involved. Popular answers included transcription of audio files and notes, changing the format of the data, entering data into statistical software, analysis, and cleaning and checking data.

**Storing data**

98% of researchers claimed to have retained the data they used to write their last research paper. Storage methods and locations were quite varied, with many researchers listing more than one type of storage (see Table 8). Only four researchers out of 83 who answered a question about how the data was stored stored no data digitally.
<table>
<thead>
<tr>
<th>Storage Type</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this storage type</th>
</tr>
</thead>
<tbody>
<tr>
<td>'A' computer</td>
<td>43</td>
<td>37.1</td>
<td>51.8</td>
</tr>
<tr>
<td>'My' computer</td>
<td>9</td>
<td>7.8</td>
<td>10.8</td>
</tr>
<tr>
<td>Secure digital storage</td>
<td>12</td>
<td>10.3</td>
<td>14.5</td>
</tr>
<tr>
<td>Shared storage or backups</td>
<td>8</td>
<td>6.9</td>
<td>9.6</td>
</tr>
<tr>
<td>USB/CD/DVD</td>
<td>12</td>
<td>10.3</td>
<td>14.5</td>
</tr>
<tr>
<td>Research Database</td>
<td>2</td>
<td>1.7</td>
<td>2.4</td>
</tr>
</tbody>
</table>

Table 8: How researchers store archival data

When asked why they kept their data, researchers gave a number of reasons, as laid out in Table 9. Encouragingly, 7 out of 83 researchers specifically mentioned potential future use by others as well as themselves.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future comparison</td>
<td>10</td>
<td>9.3</td>
<td>12.0</td>
</tr>
<tr>
<td>Future re-analysis</td>
<td>14</td>
<td>13.1</td>
<td>16.9</td>
</tr>
<tr>
<td>Other future research/publication</td>
<td>18</td>
<td>16.8</td>
<td>21.7</td>
</tr>
<tr>
<td>Future reference</td>
<td>20</td>
<td>18.7</td>
<td>24.1</td>
</tr>
<tr>
<td>Evidence or verification</td>
<td>13</td>
<td>12.1</td>
<td>15.7</td>
</tr>
<tr>
<td>Legal/ethical requirements</td>
<td>16</td>
<td>15.0</td>
<td>19.3</td>
</tr>
<tr>
<td>Data contributed to archive</td>
<td>6</td>
<td>5.6</td>
<td>7.2</td>
</tr>
<tr>
<td>Still using the data</td>
<td>6</td>
<td>5.6</td>
<td>7.2</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>3.7</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Table 9: Reasons for saving data

It is encouraging to note how many researchers listed legal, ethical and evidentiary reasons for retaining their data. It demonstrates a willingness on the part of (at minimum) this number of researchers to keep data for reasons that may be of no benefit to them. The researchers who had not kept their data gave no reason why not.

![Figure 1: Researchers' comfort with their own data storage approach](image)

Researchers were asked to rate on a 5-point Likert scale how comfortable they are with how their data is stored, with 1 being very comfortable and 5 being very uncomfortable, and give a reason for this feeling of comfort or discomfort. Researchers appear relatively comfortable, with a median
comfort rating of 2, (mean 2.2)—see also Figure 1. Reasons given for comfort include comfort in backups, a belief in the security of storage of sensitive data, and convenience. Concerns about the way data is stored include how cumbersome it can be to store data electronically and safely, concern about the lifespan of media such as DVDs and tapes, and concerns about the security of physical locations.

When asked if there was anything they would change about their data storage, 37 people (44%) said that there was; popular changes listed included assistance with backing data up, assistance with making data secure from prying eyes, help with format conversion (in some cases conversion from hard copy to digital) and additional storage space, both physically and digitally.

**Re-using data**
64 out of 85 researchers said they had re-used some data in the past, supporting their ideas about storing data for future use—future use isn’t just hypothetical, 75.3% of respondents have actually done it. Common reasons given for data re-use included further publication (or publication for different audiences), re-analysis, analysis from a different research perspective, and answering further research questions—again reflecting the reasons why researchers claimed to have kept research data.

As demonstrated in Figure 2, most researchers found re-analysis of their own data relatively simple. Many researchers commented re-use was simple because they still had access to the data, and because it was stored in such a way as to make it easy to re-analyse (for example SPSS files). Researchers who found it cumbersome to re-analyse their data cited the time taken as a reason for not finding the process easy.
**Data storage regulations**

Researchers were asked whether they had ever been required to retain data in the past. 48 researchers (56.5%) had been required to store data, 26 researchers (30.6%) had never been required to store data, and 11 researchers (12.9%) weren’t sure. The remainder of this section addresses the experiences of those 48 researchers who had definitely been required to store data.

The main type of regulation cited by researchers as having required them to store data was ethical regulation, cited by 32 researchers (66.7%). 9 researchers (18.8%) cited research funding regulations, 3 (6.3%) cited departmental policy, and 4 (8.3%) gave other types of regulations, including thesis and contractual obligations.

<table>
<thead>
<tr>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>None given</td>
<td>5</td>
<td>8.6</td>
</tr>
<tr>
<td>Data privacy/security</td>
<td>17</td>
<td>29.3</td>
</tr>
<tr>
<td>Back-ups</td>
<td>18</td>
<td>31.0</td>
</tr>
<tr>
<td>Electronic copy must be kept</td>
<td>5</td>
<td>8.6</td>
</tr>
<tr>
<td>Original hardcopy must be kept</td>
<td>3</td>
<td>5.2</td>
</tr>
<tr>
<td>External agency/software used</td>
<td>4</td>
<td>6.9</td>
</tr>
<tr>
<td>Other</td>
<td>6</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Table 10: Technical requirements on data storage

13 researchers (27.1%) were required to make the data available to other researchers in some circumstances. The circumstances under which those 13 researchers were required to make their data available varied; 7 researchers (53.8%) were required to make it available ‘upon any reasonable request’, 3 researchers (23.1%) were required to make it available at an external agency’s discretion, 2 researchers (14.4%) had to make it available to others in their team, and the final researcher only had to make data available in the form of publication.

Researchers were questioned about whether the regulations affecting their data imposed any specific technical regulations about storage; their answers are shown in Table 10.

We asked researchers how long they were required to store their data for, and the timeframes ranged from 1-10 years (see Table 11).

<table>
<thead>
<tr>
<th>1-2 years</th>
<th>3-5 years</th>
<th>6-8 years</th>
<th>8-10 years</th>
<th>&gt;10 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>3</td>
<td>24</td>
<td>18</td>
<td>2</td>
</tr>
<tr>
<td>Percentage</td>
<td>6.3</td>
<td>50.0</td>
<td>37.5</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Table 11: How long researchers were required to retain research data

Researchers were asked to rate on a 5-point Likert scale how easy it was to comply with data storage regulations. Their responses are shown in Figure 3 (overleaf).

As we can see from the Figure 3, most researchers faced with regulations about data storage found it very easy to comply with them. This general comfort is reflected in answers about what services Swinburne could provide to assist researchers in data storage requirements—researchers who answered this question almost all asked for more storage space both for physical data and digital data. Other requests included technical assistance in converting and storing digital data, back-up services and record-keeping services.
Institutional involvement

We asked respondents whether Swinburne could assist them with managing their data in any way, and if so, how. 55 researchers (64.7%) said there was nothing we could do. The answers from the remaining 30 researchers (35.3%) reflected the answers given about meeting data regulations and researchers’ concerns about their own data storage approaches, as seen in Table 11.

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digital archive space</td>
<td>10</td>
<td>30.3</td>
<td>33.3</td>
</tr>
<tr>
<td>Backup service</td>
<td>6</td>
<td>18.2</td>
<td>20.0</td>
</tr>
<tr>
<td>Data format conversion service</td>
<td>4</td>
<td>12.1</td>
<td>13.3</td>
</tr>
<tr>
<td>Physical archive space</td>
<td>3</td>
<td>9.1</td>
<td>10.0</td>
</tr>
<tr>
<td>Training</td>
<td>3</td>
<td>9.1</td>
<td>10.0</td>
</tr>
<tr>
<td>Research assistants</td>
<td>2</td>
<td>6.1</td>
<td>6.7</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>15.2</td>
<td>16.7</td>
</tr>
</tbody>
</table>

Table 12: Researcher requests for institutional assistance with data management

Researchers were also asked how comfortable they would be with ‘an institution-wide approach to data management’ (again using a 5-point Likert scale). As we can see in Figure 4, they were largely ambivalent:
Researchers were also given space to make free text comments about the idea of an institutional approach to data management, and nearly every respondent had something to say. Typical comments in support of an institutional approach included large free backup space, ease of access to the data for future research, and hopes for a more efficient work process. There were many more negative comments than positive ones, though, including concerns about the university’s ability to actually run a data archiving service, ethical concerns, and worries about where the money would come from for such a service.

### 4.5 Using data generated by others

Unfortunately, due to technical problems with the software, respondents were only given the opportunity to answer three questions about their use of data generated by others.

45.3% of all respondents said they had used data generated by others. To better understand this use, respondents were asked a free text question about how their use differed from the use of the data by researchers who collected it. Again a generative approach was taken to classifying responses, and the results are as follows:

<table>
<thead>
<tr>
<th>Results confirmation</th>
<th>New context or question</th>
<th>New analysis method</th>
<th>To generate new questions</th>
<th>Data was from an anonymous archive</th>
<th>Data generated by employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>4</td>
<td>9</td>
<td>12</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Percentage</td>
<td>11.4</td>
<td>25.7</td>
<td>34.3</td>
<td>11.4</td>
<td>8.5</td>
</tr>
</tbody>
</table>

**Table 13: Types of uses for others’ data**
It is particularly interesting to note that data is sometimes found by using an archive; this shows that researchers are willing to use anonymous data sources.

As a final question, researchers were asked on a 5 point Likert scale how likely they were to consider using another researcher’s data in the future, and given a free text space to indicate why they felt the way they did. The distribution of responses along the Likert scale are shown in Figure 5. 58 researchers (68%) gave a reason for why they thought they would or would not use other’s data in future. Typical reasons included existing research ties, expense or difficulty of obtaining data (particularly in astronomy or computer networking contexts) and an interest in drawing one’s own conclusions and confirming others’ results. Typical reasons not to use others data included ethical constraints on data sources, lack of availability of any appropriate data, concern about others’ data collection methods, and a simple lack of necessity. When researchers who have previously used others data are compared to researchers who have not previously used others’ data, they are slightly more likely to consider doing so again in the future. There is no significant difference between graduate students and other researchers in this regard.

4.6 Sharing their own data

Of 85 researchers, 30 (32.3 %) knew their data had definitely been re-used by others, 32 (37.6%) were certain no data of theirs had ever been used by others, and 23 (27.1%) weren’t sure. Those who had shared their own data were more likely to have used others’ data than those who had not shared their data.

Respondents were asked whether or not they had ever shared data with other researchers. Those who were certain they had were asked a number of further questions about that data sharing.
**With whom, and how**

The 30 respondents who had definitely shared their data were asked whom they shared it with. Results are as follows:

<table>
<thead>
<tr>
<th></th>
<th>Close colleague, same institution</th>
<th>Distant colleague, same institution</th>
<th>Unknown person, same institution</th>
<th>Close colleague, other institution</th>
<th>Distant colleague, other institution</th>
<th>Unknown person, other institution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>18</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td>60</td>
<td>3.3</td>
<td>6.7</td>
<td>10.0</td>
<td>13.3</td>
<td>6.7</td>
</tr>
</tbody>
</table>

Table 14: Who researchers share data with

As is to be expected, researchers are more likely to share data with people with whom they work closely. Looking at the ways the other researchers found the data sheds some light on how other connections are formed:

<table>
<thead>
<tr>
<th></th>
<th>Offered data</th>
<th>Research collaboration</th>
<th>Word of mouth</th>
<th>Repository or database</th>
<th>Journal publication</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td>53.3</td>
<td>13.3</td>
<td>13.3</td>
<td>3.3</td>
<td>16.7</td>
</tr>
</tbody>
</table>

Table 15: How other researchers heard about the data

Research collaborations included supervisors and students, research groups, and one inter-institution collaboration. While only one occurrence of data sharing is known to have occurred via a database or repository, it is important to note that only those who were certain their data had been re-used saw this question.

**Type of data**

Researchers who had definitely shared data were asked what type of data they had shared, and they were allowed to select more than one item from the list. Respondents generally selected fewer data types than they did when asked to describe their own data (range 1-5, median 1, mean 1.5). Results are seen in the Table 16

As we can see, the majority of data is digital (all entries in ‘Other’ were defined in free text as being digital) with only six out of 44 instances of shared data involving non-digital files (13.6%). Compare this with researchers’ own data, described in Table 5, where 38.6% of all the data described is non-digital. This indicates that sharing preference is given to digital data.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text on paper</td>
<td>5</td>
<td>11.4</td>
<td>16.7</td>
</tr>
<tr>
<td>Numerical data on paper</td>
<td>1</td>
<td>2.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Non-digital images</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Digital text</td>
<td>9</td>
<td>20.5</td>
<td>30.0</td>
</tr>
<tr>
<td>Spreadsheet/digital numerical data</td>
<td>11</td>
<td>25.0</td>
<td>26.7</td>
</tr>
<tr>
<td>Databases</td>
<td>7</td>
<td>15.9</td>
<td>23.3</td>
</tr>
<tr>
<td>Digital images</td>
<td>5</td>
<td>11.4</td>
<td>16.7</td>
</tr>
<tr>
<td>Audio on tape</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Digital audio</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Video on tape</td>
<td>1</td>
<td>2.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Digital video</td>
<td>0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td>9.0</td>
<td>13.3</td>
</tr>
</tbody>
</table>

Table 16: Type of data respondents shared with other researchers
Restrictions on data use

All those who had shared data were asked if they had placed any restrictions on its use; only seven had. Of those, two had asked the recipients of the data to abide by ethics/confidentiality restrictions, two had asked for some acknowledgement on publication, one had asked to see the final version of any publication, one had asked for acknowledgement only if he or she was an author, and one didn’t describe the restrictions.

Likelihood of sharing data in the future

Participants were asked to rate how likely they were to share their data in the future on a 5 point Likert scale, where 1 represented highly likely, and 5 represented highly unlikely. The distribution of responses can be seen in Figure 5. When comparing Figures 1 and 2 it may look as though researchers are more likely to share their own data than to use others’; however this difference is not significant. There is a moderate correlation ($r=0.455$) between willingness to share data, and willingness to use others’ data.

Again, respondents were given the opportunity to say why they felt the way they did about sharing their data. Common reasons for sharing data included cultural “that’s how we do it in my field and I have used others’ data”-type reasons, and simple lack of reservation. Concerns about sharing data included the complexity of the data involved (and the fear of misinterpreted results), ethical concerns, and a desire for acknowledgement in publication. Again, there is no significant difference between students and other researchers in likelihood of sharing data in the future.
4.7 The New ARC rules

Given that this survey was run primarily to assess how we can help Swinburne researchers meet their obligations under the new ARC rules about storing data, our first question was how likely researchers were to deposit their data in a repository under the new rules.

Table 17: Perceived advantages of the new ARC rules

<table>
<thead>
<tr>
<th>Advantage</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this advantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data preservation</td>
<td>6</td>
<td>7.3</td>
<td>9.3</td>
</tr>
<tr>
<td>Data back-up</td>
<td>9</td>
<td>11.0</td>
<td>13.9</td>
</tr>
<tr>
<td>Security of data</td>
<td>9</td>
<td>11.0</td>
<td>13.9</td>
</tr>
<tr>
<td>Access to others’ data</td>
<td>7</td>
<td>8.5</td>
<td>10.8</td>
</tr>
<tr>
<td>Sharing own data with others</td>
<td>20</td>
<td>24.4</td>
<td>30.9</td>
</tr>
<tr>
<td>Cultural shift toward sharing and community</td>
<td>7</td>
<td>8.5</td>
<td>10.8</td>
</tr>
<tr>
<td>Good publicity</td>
<td>3</td>
<td>3.7</td>
<td>4.6</td>
</tr>
<tr>
<td>Good use of resources</td>
<td>4</td>
<td>4.9</td>
<td>6.2</td>
</tr>
<tr>
<td>Evidence/verification of results</td>
<td>9</td>
<td>11.0</td>
<td>13.9</td>
</tr>
<tr>
<td>‘None for me’</td>
<td>5</td>
<td>6.1</td>
<td>7.7</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>3.7</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Common reasons why researchers would deposit their data included “because it’s the rules”, safety and security of the data in a repository, and a genuine willingness to share the data with others. The main point against depositing data in a repository was ethical obligation; researchers were quite concerned that human participants should remain anonymous. Researchers were also sceptical that depositing data would be a useful exercise, and even worried that a repository may not have enough space for their terabytes-large data sets.
79 researchers responded to a question about the benefits to them of data sharing, and several themes emerged in their responses as documented in Table 17.

Despite being able to see some advantages in depositing data to a repository, researchers also perceived many disadvantages in the proposal (74 respondents commented). It is clear from the raw responses that researchers take their ethics and confidentiality obligations very seriously, which is a good thing. They are also concerned about loss of control of their data, worried about its misuse or theft and concerned about their own intellectual property (see Table 18 below):

<table>
<thead>
<tr>
<th>Concern</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this concern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidentiality/ethics</td>
<td>24</td>
<td>34.3</td>
<td>32.4</td>
</tr>
<tr>
<td>Misuse or theft of data</td>
<td>22</td>
<td>31.4</td>
<td>29.7</td>
</tr>
<tr>
<td>Concern that their work be acknowledged</td>
<td>2</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Time uploading takes</td>
<td>9</td>
<td>12.9</td>
<td>12.2</td>
</tr>
<tr>
<td>Cost of repository</td>
<td>5</td>
<td>7.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Usability of repository</td>
<td>3</td>
<td>4.3</td>
<td>4.1</td>
</tr>
<tr>
<td>Reliability of repository</td>
<td>5</td>
<td>7.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Bureaucracy</td>
<td>5</td>
<td>7.1</td>
<td>6.8</td>
</tr>
<tr>
<td>Duplication of existing discipline services</td>
<td>2</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>‘None’</td>
<td>2</td>
<td>2.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>17.1</td>
<td>16.2</td>
</tr>
</tbody>
</table>

**Table 18: Perceived disadvantages of the ARC proposal**

Interestingly, one participant was quite adamant that research culture would be destroyed by data sharing, in sharp contrast to those above who thought research culture would be improved by it. Clearly data sharing is an emotive issue and needs to be approached carefully with researchers.

<table>
<thead>
<tr>
<th>Restriction</th>
<th>Number</th>
<th>Percentage of total responses</th>
<th>Percentage of respondents listing this restriction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control over who used the data</td>
<td>27</td>
<td>33.8</td>
<td>36.0</td>
</tr>
<tr>
<td>Ethics obligations upheld</td>
<td>15</td>
<td>18.8</td>
<td>20.0</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>12</td>
<td>15.0</td>
<td>16.0</td>
</tr>
<tr>
<td>Intellectual property protection</td>
<td>10</td>
<td>12.5</td>
<td>13.3</td>
</tr>
<tr>
<td>Notification when data is used</td>
<td>6</td>
<td>7.5</td>
<td>8.0</td>
</tr>
<tr>
<td>To know how the data will be used</td>
<td>5</td>
<td>6.3</td>
<td>6.7</td>
</tr>
<tr>
<td>Data security</td>
<td>5</td>
<td>6.3</td>
<td>6.7</td>
</tr>
<tr>
<td>Access to the data themselves</td>
<td>5</td>
<td>6.3</td>
<td>6.7</td>
</tr>
<tr>
<td>Non-commercial share-alike as in Creative Commons</td>
<td>3</td>
<td>3.8</td>
<td>4.0</td>
</tr>
<tr>
<td>No restrictions</td>
<td>3</td>
<td>3.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Other</td>
<td>9</td>
<td>11.3</td>
<td>12.0</td>
</tr>
</tbody>
</table>

**Table 19: Restrictions researchers would place on data**

Finally, we asked researchers what conditions, if any, they would like to place on data they had stored in a repository. 75 researchers answered this question, and their answers generally reflected the perceived disadvantages of repository data storage.

5. **Discussion and conclusions**

The Swinburne survey demonstrated that Swinburne researchers are not so very different from their counterparts at larger institutions, and graduate students are not any different from any other type of researcher.
5.1 Existing data practices
It is readily apparent that researchers are already used to retaining their research data for a wide variety of purposes, and that many of them re-use their own data. This implies that research culture includes an understanding of the importance of data as a commodity.

Researchers are also relatively used to sharing data, either by using data generated by others, or by sharing their own. They are also amenable to engaging in either of these practices in the future, however the vast majority of sharing goes on among people who know each other, and thus in an established relationship of trust. The idea of sharing data with strangers is relatively new, and the majority of researchers surveyed could see disadvantages in such a situation, though they were not necessarily averse to it.

It is interesting to note that shared data is more likely to be digital than personal data. This suggests that data that exists in digital form is more likely to be shared. If this implication is borne out, assistance with digitising data and encouragement and training to create data in a digital form may well be useful ways Swinburne could encourage data sharing.

5.2 Concerns about sharing data
There are two themes that run through researchers’ concerns about sharing their research data: that ethical and confidentiality obligations are observed, and that they maintain control of their data. These concerns reflect the issues raised in the DART project [13].

Researchers are very aware of the ethical obligations to which their data is subject, but the raw text of their comments suggest that this is a concern that goes deeper than just following the rules. Many researchers spoke of the trust and relationships they built with their participants, and how they were concerned that any data sharing obligations or institutional involvement in data practices may result in a breach of this trust. Any institutional data sharing approach will have to take into account the rightful concerns of researchers about their ethical obligations.

The data control issue is a trickier one to address. Many researchers simply did not want any other researcher to have access to their data under any circumstances. Many had a less hard-line approach, but wanted varying degrees of control varying from notification to the right to determine exactly who might have access to their data and for what purpose. Reasons for this concern emerged in raw comments; researchers would like acknowledgement for the work they do in collecting the data, but they are also concerned about being beaten to publication or their data being in some way stolen; ownership was seen as very important. Given that individual endeavour and acknowledgement are key to advancement in today’s research climate, particularly with funding tied to research assessments, these concerns must be addressed when approaching compliance with the new ARC rules.

There is a third current of dissent to the ARC rules, which was not so clear as in the DART work [13], but was nevertheless there: researchers are concerned that the system will be time consuming to use and subject them to unnecessary bureaucracy, thus taking them away from their research.

5.3 Perceived advantages in data sharing
Researchers’ articulation of the advantages of data sharing was neither as forceful nor as common as their articulation of their concerns. While some advantage was seen in being able to access others’ data, many researchers said that they did not know whether it would be that useful to be able to do so. There is a certain amount of recognition that data sharing might be good for research culture, but this is fairly muted as well. If any data sharing initiative outside researchers’ own
networks is to be successful, the benefits of the initiative to the individual researcher will have to be clearly articulated, or buy-in is likely to be low.

5.4 Assistance Swinburne could provide now
Despite being relatively happy with how they managed their data, researchers also saw a number of ways Swinburne could assist them, in particular in provision of space.

Researchers were very interested in the provision of both physical and digital spaces to store their data, provided certain conditions were met. If Swinburne was to provide successful physical and digital storage spaces, those spaces would need to be secure, and provide each researcher with access to their own data and give them confidence that there would be no unauthorised access to their data. Researchers are also concerned about backing up their data, so digital storage would have to be well backed up and secure, but nonetheless accessible to authorised users of any given data set.

Another service researchers mentioned as being particularly useful is digitisation assistance, both with technical issues and in terms of the actual work to be done. Researchers recognise that effective digitisation can be difficult and time consuming, but that it is also useful. Given that undigitised data is shared less than digitised data, assistance with digitisation is also conducive to data sharing.

Some researchers mentioned that they were interested in training in data management, but until we have a better understanding of what data management is going to entail under the new ARC rules, and what assistance we can provide in data management, training is unlikely to be very useful.

5.5 Conclusions
There is scope for Swinburne to provide institutional assistance with data management now, in the form of secure physical and digital space for data storage and assistance with digitisation. At this stage, however, use of such spaces should remain optional.

Researchers understand very clearly the value of their data, and are interested in ownership and protecting the confidentiality of human data results; thus they are concerned about the imposition of regulations about what should happen to their data, particularly when it comes to mandated sharing. However, when data is under researchers’ own control, they are not averse to sharing it, nor to using data generated by others. It is the blunt nature of institutional rules (and the potential to become bound up in bureaucracy and technical software problems) that they find worrisome.

Researchers have a strong investment in their own data, and could resist what they might perceive to be interference in their management of it. Having said that, a measured approach, respect for researchers’ concerns about data ownership, control and confidentiality, a usable system (and rules that researchers can follow), and clear examples of the benefits researchers might gain will go a long way towards encouraging data sharing in a more anonymous context.

6. Acknowledgements
The work to prepare this review was funded by the ARROW Project (Australian Research Repositories Online to the World, www.arrow.edu.au). The ARROW Project is funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education.
7. References


Appendix A: The survey
Overleaf, you will find a copy of the survey on which this report is based. The survey is presented in flowchart form. The symbols in the flowchart are:

- Multi-line text box. The size of the box gives some indication of the number of lines
- Single line text box, for entering short sentences
- Radio button, where participants may only select a single option of those available

Re-using the survey
This document, including the survey, is licensed under the Creative Commons Attribution-Noncommercial-Share Alike License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/2.5/au/. If you do re-use the survey, we would be interested in hearing about your experiences with it, so please get in touch with dmckay@swin.edu.au.
About the survey

The Australian Research Council (ARC) has changed its funding rules so that, beginning in 2008, recipients of ARC funding will be required either to undertake to store their research data in a repository, or provide reasons why they will not do so.

The information you provide in this survey will help the Information Resources department (the people who are responsible for the library, the Swinburne web site, and copyright) at Swinburne University of Technology understand how local researchers are using, sharing, and storing data.

Similar studies have recently been done at both Melbourne University and Monash, and Swinburne is doing this study to gain an understanding of the specific needs of a smaller university with diverse research interests. In the longer term the information generated by this survey may be used to provide assistance in managing research data at an institutional level, and to aid researchers in meeting ARC requirements.

For the purposes of this survey (these are not the ARC’s criteria; those are presently unknown), 'data' is any material you analyse for the purpose of research (a good guide is whether your analysis is likely to go into a thesis or publication). This means that data can include handwritten notes about an interview, photographs, sculpture or film, scientific observations, computer program output, etc. Please answer the survey with your data in mind.

This survey will take approximately 20 minutes to complete. Your responses are completely anonymous, and you will not be personally identifiable in any material published about the results. If you have any questions or comments about the survey, please feel free to contact the survey author (Dana McKay) at dmckay at swin dot edu dot au.

About you

What is your primary research position?
- Masters student
- Doctoral student
- Postdoctoral fellow
- Research assistant/research fellow
- Academic (lecturer/senior lecturer/professor)
- Consultant
- Medical researcher

Are you completing this survey on your own behalf, or on behalf of your research group?
- On my own behalf
- On behalf of my research group

How long have you been in a research role?
- 0-2 years
- 2-5 years
- 5-10 years
- 10-20 years
- 20+ years

How many people in your research group (including yourself)?

Approximately how many years has your research group been running (in any form, even if it was an informal arrangement)?
- Fewer than 6 months
- 6 months to 1 year
- 1-2 years
- 3-5 years
- 5-10 years
- Longer than 10 years (how long?)

How long have you been a member of your research group? (If less than 1 year, please enter a decimal approximation)

What field(s) is your research in?

____________________________________________________________________
Please describe the type of data you typically work with (e.g. survey responses, scientific observations, primary source materials, notes from focus groups or interviews):

What format(s) was the data in when you got it?
- Text on paper (including handwritten notes)
- Numerical data on paper
- Images on paper or other non-electronic medium
- Digital text (e.g. Word documents, text documents)
- Spreadsheet (e.g. Excel documents)
- Database (e.g. Access)
- Digital images (e.g. jpegs)
- Audio on tape
- Digital audio (e.g. MP3s)
- Video on tape
- Digital video (e.g. DVDs, mpegs)
- Other (please describe below)

Is there anything else that distinguishes your data from that of your colleagues, or from the data used in other disciplines (for example cultural issues, ethics requirements)?
- Yes (please explain below)
- No

Using data generated by other researchers

Is there anything else that distinguishes your data from that of your colleagues, or from the data used in other disciplines (for example cultural issues, ethics requirements)?
- Yes
- No

When answering the next few questions please think about the last time you used data generated by someone else

How did you find the data?
- Given by a colleague
- Internet search
- Data repository or database
- Other (please describe below)

Who created/gathered the data?
- A close colleague from the same institution as you
- Someone from the same institution as you with whom you had had limited contact
- Someone at the same institution with whom you had had no prior contact
- A close colleague at another institution
- Someone at another institution with whom you had had limited contact
- Someone at another institution with whom you had had no prior contact
- Unknown (data is from a research database, for example)
Were there any restrictions or conditions placed on your use of the data?
- Yes (please explain below)
- No

What format(s) was the data in when you got it?
- Text on paper (including handwritten notes)
- Numerical data on paper
- Images on paper or other non-electronic medium
- Digital text (e.g. Word documents, text documents)
- Spreadsheet (e.g. Excel documents)
- Database (e.g. Access)
- Digital images (e.g. jpegs)
- Audio on tape
- Digital audio (e.g. MP3s)
- Video on tape
- Digital video (e.g. DVDs, mpegs)
- Other (please describe below)

How different was your use of the data from the original researcher's use of the data (think about things like research questions, analysis methods)?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>No different at all</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Please comment a little on the differences between how you used the data and how the original researcher used it:

How easy was the data to obtain?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very easy</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

How sure were you about the provenance of the data (that the data was real and had not been interfered with in any way)?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very sure</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

15. How old was the data?
- 0-6 months
- 6 months-1 year
- 1-2 years
- 3-5 years
- 5-10 years
- Older than 10 years (please specify)
How useful was the data?

1  2  3  4  5
Very useful O O O O O  Not at all useful

Why was it useful/not useful?

How easy was the data to use, compared with generating your own data?

1  2  3  4  5
Much easier than generating my own data O O O O O
Much harder than generating my own data

How much work did it take to get the data to a point where you could use it (e.g. cleaning the data, entering it into a computer, changing the format of the data)?

1  2  3  4  5
No work at all O O O O O
A lot of work

How long did it take to get the data into a format where you could use it?
- Less than a day
- 1-2 days
- About a week
- About a month
- Longer (please specify) ___________

What did you have to do to the data to get it to a point where you could use it?

Please say why you chose to use this data:

How likely would you be to use data generated by another researcher in the future?

1  2  3  4  5
Very likely O O O O O  Very unlikely

Managing your own data

When answering the next few questions, please think about data you used in the last research report or paper you created about data you generated or collected

Please describe the data you used:
When the data was created, was it created in a computer-readable format (e.g., notes taken on a PDA, astronomical data created in a digital format, digital photos), or a non-computer-readable format (e.g., handwritten notes, audio or video recorded on tape, paper seismic graphs)?
- All data was computer readable
- No data was computer readable
- A combination of the above (please describe below)

Did you have to computerize the data, or change its format (e.g., transcribing audio, forcing raw text into a spreadsheet) before you could analyse it for your paper?
- Yes
- Some of it (approximately what proportion) _____
- No

How long did you spend reformatting the data?
- Less than an hour
- 1-2 hours
- About a day
- About a week
- About a month
- More than a month (how long) _________

What did you have to do to the data to get it into a form where you could analyse it?

Have you kept the data in some form (e.g., handwritten notes in a box, files on a computer, etc.)?
- Yes
- No

How is the data stored?

Why didn’t you keep the data?

Why did you keep the data?

How comfortable are you with the way the data is stored?

Why are you comfortable/uncomfortable?

In an ideal world, is there anything you would change about the way the data is stored?
- Yes
- No
What would you change about the way the data is stored?

For any data you have ever generated, have you been obliged by regulations (for example funding regulations, ethical consent procedures, departmental policy) to store that data?

Yes  Not that I am aware of  No

Please answer the next few questions thinking about the last time you were required by regulations to store data

What kind of regulations imposed the obligation to store the data?
- Research funding regulations
- Ethical consent regulations
- Departmental policy
- Other (please state) __________

How long were you required to store the data for?
- Less than 1 year
- 1-2 years
- 3-5 years
- 6-8 years
- 8-10 years
- More than 10 years (how long) __________

Were you required to make the data available to other researchers?

Yes  No

Under what circumstances were you required to make the data available to other researchers?
- At your discretion
- Upon any reasonable request
- At the discretion of an external agency (e.g. the research funder)
- Under any and all circumstances (e.g. open access in a database)
- Other (please describe below) __________

What data storage requirements (e.g. making the data available over the internet, backing the data up, putting the data in a database) did you have to meet?

How easy was it to comply with the regulations about this data?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very easy</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Please describe any services that would have helped you comply with the regulations about this data.

Are there any services you can think of that would help you manage your research data in general?
- Yes (please describe below)
- No

How comfortable would you be with an institution-wide approach to data storage and management?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Very uncomfortable</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
</tbody>
</table>

Reusing your data

Have you ever re-analysed your own data to answer new research questions?
- Yes
- No

Please describe some of the differences between the original research question and the new one:

In the instance you just described, how difficult was it to re-analyse your own data for new purposes?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Very difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
</tr>
</tbody>
</table>

Why was it easy/difficult?

Has research data you generated ever been used by anyone else?
- Yes
- No
- Don’t know

When answering the next few questions please think about the last time someone else used data you generated

Was that person:
- A close colleague from your institution
- Someone from your institution with whom you had had limited contact
- Someone from your institution with whom you had had no prior contact
- A close colleague from another institution
- Someone from another institution with whom you had had limited contact
- Someone from another institution with whom you had had no prior contact
- Don’t know (for example because they downloaded the data anonymously from a database)
How did that person find out about your data?
- You offered it to them
- They heard about it through word of mouth
- They found out about it on the internet
- They found it in a data repository or database
- Other (please state) _______________

How old was the data?
- 0-6 months
- 6 months-1 year
- 1-2 years
- 3-5 years
- 5-10 years
- Older than 10 years (please specify) _______________

Did you place any conditions on the use of your data?
- Yes
- No

Please describe the conditions you placed on the use of your data:

What format(s) did the recipient get the data in?
- Text on paper
- Numerical data on paper
- Images on paper or other non-electronic medium
- Digital text (e.g. Word documents, text documents)
- Spreadsheet (e.g. Excel documents)
- Database (e.g. Access)
- Digital images (e.g. jpegs)
- Audio on tape
- Digital audio (e.g. MP3s)
- Video on tape
- Digital video (e.g. DVDs, mpegs)
- Other (please describe below)

Please say a little bit about the data (what it was, what the recipient intended to use it for, etc).

How likely would you be to use data generated by another researcher in the future?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very likely</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Very unlikely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Why/why not?


You may or may not know that the ARC from 2008 will require funding recipients to deposit data in an online repository, or supply a reason why they will not do so. How likely are you to put your data in a repository under such a regime?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very likely</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Please give some of the reasons why you would deposit your data:

Why are you unlikely to deposit your data?

What three advantages can you see in such a proposal?

Outline three problems you see with such a proposal:

What difficulties would you personally face if you were required to deposit research data into a repository?

If you personally were required to deposit research data to a repository, what conditions would you wish to place on its use?

Please leave any other comments you may have any other aspects of research data management and sharing here:

Contact details

If you are prepared to be contacted to answer further questions, or if you want to go on the draw to win the iPod, please leave your email here.

Please note that your email address will not be distributed or disseminated in any way, nor will it be associated with your responses to the survey, which are completely anonymous. Your email address will only be used for the purposes for which you explicitly give permission below, and for no other purpose.

If you have any questions or would like to contact me (Dana) write to dmckay at swin dot edu dot au.

- Yes, I would like to be in the draw to win the iPod (you must use a real email address or we will have no way to contact you)
- Yes, I am open to being contacted for further questions

Your email:

Thank you for taking the time to complete this survey. If you have any questions about the survey, please feel free to contact the survey author (Dana McKay) at dmckay at swin dot edu dot au.