Research data management requirements online survey report

iridium project output

Project Information			
Project Title	*iridium - Institutional Research Data Management at Newcastle University*		
Project Hashtag	*#iridiummrd*		
Start Date	1 October 2011	End Date	31 March 2013
Lead Institution	Newcastle University		
Project Director	Janet Wheeler		
Project Manager	Lindsay Wood		
Contact email	lindsay.wood@ncl.ac.uk		
Project Web URL	http://research.ncl.ac.uk/iridium/		
Programme Name	*JISCMRD02*		
Programme Manager	Simon Hodson		
Document Information			
Author(s)	Dr Simon Kometa		
Project Role(s)	Research Computing Specialist (ISS)		
Date	17 August 2012	Filename	*iridium*_online_survey_report_17_8_2012_v2.1_SK.docx
URL	http://research.ncl.ac.uk/iridium/outputs/		
Access	DRAFT for external review		

Document History		
Version	**Date**	**Comments**
v2.1	17 August 2012	*iridium*_online_survey_report_17_8_2012_v2.1_SK.docx by SK

NEWCASTLE UNIVERSITY

Iridium

Research Data Management

Online Requirement Survey

Contents

Iridium Online Survey

RDM Requirement Analysis

Executive Summary

Key Findings

- One hundred and twenty eight projects completed the online survey and over half of the projects are from the Faculty of Medical Sciences (nearly 52%).
- Over 97% of projects' data is in digital / physical format.
- Generally projects have many files e.g. 23.4% of projects have between 100- 1000 files, 28.1% between 10-100 files and 19.5% have between 1000 – 10000 files.
- Thirty one percent of projects' files take up to 4GB of space. Just over 11% of projects' files take up between 64 – 0.5TB. One project had more than 100TB but no project required more than 1 PB.
- For nearly 54% of projects space required at collection is greatly different from space required after processing / analysis. However for nearly 29% of projects the space was not greatly different.
- There was lots of variability in data file format amongst projects. Some projects used many different file formats indeed. However Excel was the most common file format.
- About 30% of projects said they store their data on ISS managed systems; about 18% of projects used academic unit managed systems and about 21% used personal systems. A small number of projects used external systems / services such as cloud and SurveyMonkey.
- The majority of projects (nearly 43%) intend to keep data for 5 to 10 years. Just over 29% intend to keep data for 10 – 25 years and about 17% intend to keep data for more than 25 years. Eleven percent intends to keep data for 1 – 5 years.
- 93% of projects have multiple copies / partial back up of data.
- Nearly 49% of the projects have tested how successful it will be to retrieve backed up data.
- Just over 73% of projects share their data with others within the University.
- Just over 50% of projects share data externally.
- Of 64 projects (50%) who share data external to the University nearly 30% don't have any agreement in place while nearly 33% have *other* types of agreement not specified.
- About 56% of projects have a data management plan or partial / informal plan.
- Nearly 65% of projects don't have any specific tool for RDM.
- The majority (46%) of projects said that they do not have any deletion policy and just fewer than 17% have a deletion policy.
- For over 75% of projects data have to be quite secure or very secure.
- 84% of projects said they store their data quite securely or very securely.

- Over 90% of projects used password, anonymisation or physical measures for data security.
- Not many projects are aware of the policies and legislation that applies to their data e.g. only 63% of projects are aware of DPA and only 40% are aware of FOI.
- 64% of projects said that the PI should have the primary responsibility for RDM support. Next in line was the Research Associate with just 11%.
- 55% of projects believe that going forward the PI should still have the primary responsibility for RDM support. All the other officers got less than 10% of the vote except for computing support officer where just over 17% of projects think they should have the primary responsibility for RDM support.
- Only 5 projects said they are aware of training sessions and materials on RDM.
- 60% of projects gave a positive response to make their research data publicly available at the end of the project.
- 73% of projects have not deposited any of their data in a data repository.
- Nearly 60% of projects are willing to submit data to a data repository.
- An overwhelming majority of projects (nearly 80%) are happy to submit data to a repository at the publication stage.
- Nearly 41% of projects are willing to make data supporting any publication available immediately.
- Nearly 73% of projects are willing to share data if they have control over who can access the data.
- There is no clear consensus from projects on intellectual property rights (IPR), just over 30% of projects believe that it is owned by the University and about 17% of projects do not know; just over 19% think it belongs to their research group and 10.4% said *other*. For about 7% and 15% of projects it belonged to the funder and the researcher respectively.
- The majority of project was funded by either charity or research council 35.5% and 31.4% respectively, that is, a total of nearly 67%.

Introduction

This is the final report of the findings of the Iridium Online Survey on Research Data Management (RDM) that aims to solicit requirements for RDM. A total of 128 projects across the three faculties completed the online survey.

1 Context of Responses

1.1 Name of Research Project

Some projects had many projects and a variety of names.

1.2 Faculty

Over half of projects are from the Faculty of Medical Sciences (51.6%) followed by SAgE with 26.6% and then HaSS with 17.2%. Six projects (4.7%) did not specify their faculty (Table 1.2).

Table 1.2: Projects by faculty

Faculty		Frequency	Percent
Valid	HaSS	22	17.2
	FMS	66	51.6
	SAgE	34	26.6
	Total	122	95.3
Missing	System	6	4.7
Total		128	100.0

2 Thinking About Your Data

2.1 Format of Data

- **What format are your data in?**

The majority of projects' data is in either digital or physical format (56.3%) while 41.4% store data only digitally and just below 1% only physically. Cumulatively nearly 99% of projects store data as digital, physical or both (table 2.1a).

Table 2.1a: Data formats

Data		Frequency	Percent	Cumulative Percent
Valid	Both	72	56.3	56.3
	Digital	53	41.4	97.7
	Physical	1	.8	98.5
	Other (please specify)	2	1.6	100.0
	Total	128	100.0	

The distribution of data format by faculty is shown on table 2.1b below. There is a statistically significant association between faculty and data storage format with a chi-square value of 9.07 and p=0.011 (<0.05). A higher proportion of projects from FMS and HaSS are more careful with their data as over 68% and over 66% respectively store data in both digital and physical formats. Just over 36% of projects in SAgE store data digitally and physically. A higher proportion (63.6%) of projects from SAgE store data digitally only compare to projects in HaSS (31.8%) and FMS (33.8%). Storing data both digitally and physically provide an extra level of safety as the demise of one media is possible and one can rely on the other.

Table 2.1b: Cross tabulation of faculty and data format

			What format are your data in?		
			Both	Digital	Total
Faculty	HaSS	Count	15	7	22
		% within Faculty	68.2%	31.8%	100.0%
	FMS	Count	43	22	65
		% within Faculty	66.2%	33.8%	100.0%
	SAgE	Count	12	21	33
		% within Faculty	36.4%	63.6%	100.0%
Total		Count	70	50	120
		% within Faculty	58.3%	41.7%	100.0%

Chi-Square 9.07, df=2 p=0.011

3 For Any Digital Research Data

3.1 Approximate Number of files

- **Approximately how many files exist?**

The number of files per project is very variable, 28.1% of projects have between 10- 100 files, 23.4% have between 100-1000 files and 19.5% have between 1000 – 10000 files. It is worth noting that over 9% have between 10000 – 100000 files. Another 9% have between 100000 – 1000000 files and at the opposite end about 9% have 1- 10 files (table 3.1a).

Table 3.1a: Number of files per projects

Number of Files		Frequency	Percent
Valid	1 - 10	11	8.6
	10 - 100	36	28.1
	100 - 1000	30	23.4
	1000 - 10000	25	19.5
	10000 - 100000	12	9.4
	100000 - 1000000	12	9.4
	Total	126	98.4
Missing	System	2	1.6
Total		128	100.0

The distribution of number of files owned by projects by faculty is shown on table 3.1b below. There is also a statistically significant association between number of files and faculty with a chi-square value of 20.38 and p = 0.027 (<0.05). In HaSS over one fifth of projects have between 1 – 10 files. The corresponding proportions for FMS and SAgE are only 7.7% and 3% respectively. The proportions of projects with large number of files (100000 – 1000000) are similar for HaSS and FMS with over 9% respectively. The corresponding proportion for SAgE is slightly higher at just over 12%.

Table 3.1b: Cross tabulation of faculty and number of files

			Approximately how many files exist?						Total
			1 - 10	10 - 100	100 - 1000	1000 - 10000	10000 - 100000	100000 - 1000000	
Faculty	HaSS	Count	5	11	2	2	0	2	22
		% within Faculty	22.7%	50.0%	9.1%	9.1%	.0%	9.1%	100.0%
	FMS	Count	5	15	20	13	6	6	65
		% within Faculty	7.7%	23.1%	30.8%	20.0%	9.2%	9.2%	100.0%
	SAgE	Count	1	9	5	9	5	4	33
		% within Faculty	3.0%	27.3%	15.2%	27.3%	15.2%	12.1%	100.0%
Total		Count	11	35	27	24	11	12	120
		% within Faculty	9.2%	29.2%	22.5%	20.0%	9.2%	10.0%	100.0%

Chi-Square 20.28, df=10 p=0.027

3.2 Total Amount of Space taken by Files

- **What is the total amount of space your files take up at this stage?**

The highest 31% of projects' files take up to 4GB of space. Next highest is just over 15% of projects with space of between 4 - 16 GB. About 14% of projects' files needed between 16 – 64GB and between 1TB – 10TB respectively. Just over 11% of projects' files take up between 64 – 0.5TB. No project said that the file space was more than 1 PB. These percentages are shown on figure 3.2 below.

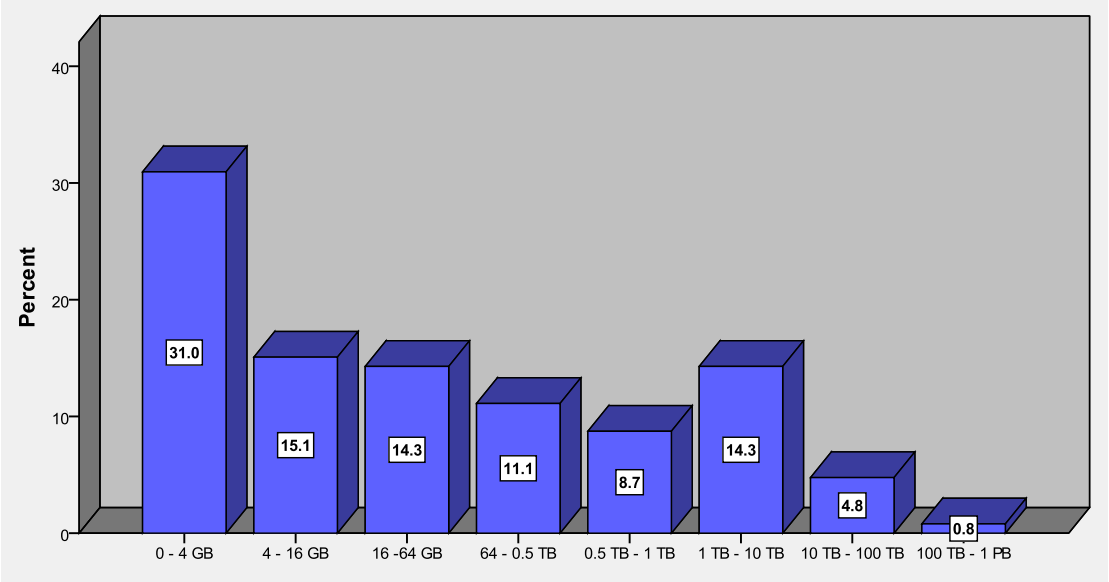

Figure 3.2: Total amount of space taken by files

The distribution of storage space by faculty is shown on table 3.2 below. There is no statistically significant association between space and faculty with chi-square value of 18.73 and p = 0.175 (>0.05). That is, there is no difference in space requirements across faculty. However, higher proportion (50%) of projects in HaSS used up to 4GB of space. The corresponding percent for FMS and SAgE are only 29% and 21% respectively.

Table 3.2: Cross tabulation of faculty and total amount of space taken by files

			What is the total amount of space your files take up at this stage?								Total
			0 - 4 GB	4 - 16 GB	16 -64 GB	64 - 0.5 TB	0.5 TB - 1 TB	1 TB - 10 TB	10 TB - 100 TB	100 TB - 1 PB	
Faculty	HaSS	Count	11	4	3	3	0	0	1	0	22
		% within Faculty	50.0%	18.2%	13.6%	13.6%	.0%	.0%	4.5%	.0%	100.0%
	FMS	Count	19	11	11	7	4	11	1	1	65
		% within Faculty	29.2%	16.9%	16.9%	10.8%	6.2%	16.9%	1.5%	1.5%	100.0%
	SAgE	Count	7	4	4	3	7	6	2	0	33
		% within Faculty	21.2%	12.1%	12.1%	9.1%	21.2%	18.2%	6.1%	.0%	100.0%
Total		Count	37	19	18	13	11	17	4	1	120
		% within Faculty	30.8%	15.8%	15.0%	10.8%	9.2%	14.2%	3.3%	.8%	100.0%

Chi-Square 18.73, df=10 p=0.175

3.3 Total Amount of Space required at Collection

- **Does the total amount of space required at collection differ greatly after processing/analysis?**

For nearly 54% of projects the space required at collection is greatly different from the space required after processing / analysis. For about 29% of projects it is not. Nearly 18% of projects did not know. These percentages are shown on the pie chart below (figure 3.3).

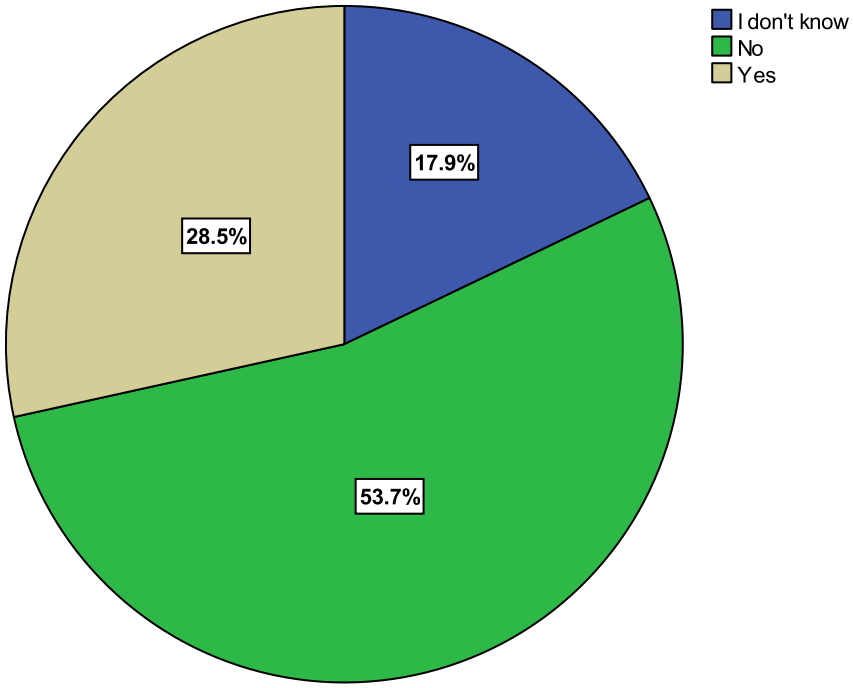

Figure 3.3: Total amount of space required at collection

The result of comparing space required at collection by faculty is shown on table 3.3 below. There is no statistical significant association of space required at collection with faculty with a chi-square value of 5.70 and p=0.223 (>0.05). Space required is similar across faculty.

Table 3.3: Cross tabulation of faculty and amount of space required at collection

			Does the total amount of space required at collection differ greatly after processing/analysis phase?			
			I don't know	No	Yes	Total
Faculty	HaSS	Count	3	15	3	21
		% within Faculty	14.3%	71.4%	14.3%	100.0%
	FMS	Count	14	31	20	65
		% within Faculty	21.5%	47.7%	30.8%	100.0%
	SAgE	Count	3	19	11	33
		% within Faculty	9.1%	57.6%	33.3%	100.0%
Total		Count	20	65	34	119
		% within Faculty	16.8%	54.6%	28.6%	100.0%

Chi-Square 5.70, df=4 p=0.223

3.4 Top Three Data Formats

- **What are the top three formats that your data are stored in?**

There was lots of variability in this question not a single format was common amongst projects. Indeed some projects have many different formats of files. However Excel was the most common file format.

3.5 Data Storage Location

- **Where do you store your data?**

This was a multiply response question where projects were allowed to select all that apply to them. The total number of selection made across all the available storage location/systems was 273. Note that 128 projects completed the online survey and for this question we had 127 valid responses. Out of the 273 selections made 81 of them (nearly 30%) said they store their data on ISS managed systems; 17.6% of projects used academic unit managed systems and nearly 21% used personal systems. A small number of projects used external systems / services such as cloud and SurveyMonkey. The percent of projects who selected the ISS managed system made up nearly 64% of the total (i.e. 81/127). See table 3.5a for details.

Table 3.5a: Data storage location

Storage Location	Responses		Percent of Cases
	Frequency	Percent	
ISS managed systems	81	29.7	63.8
Academic Unit managed systems	48	17.6	37.8
Project managed systems	23	8.4	18.1
Personal systems	57	20.9	44.9
Other (please specify)	13	4.8	10.2
Off-campus managed system	11	4.0	8.7
External Cloud (IaaS) e.g. Amazon/EC2	8	2.9	6.3
External Service (SaaS) e.g. SurveyMonkey	3	1.1	2.4
At home	29	10.6	22.8
Total	273	100.0	215.0

The distribution of storage location of projects' data by faculty is shown on table 3.5b below. There was no statistically significant association between data storage location and faculty however a higher proportion of projects from FMS seem to be using the data storage location more than projects from HaSS or SAgE. For example for ISS managed systems the percentages are HaSS 17.9%, FMS 51.3% and SAgE 30.8%. This pattern is repeated across all the other storage location except for External Cloud where projects in SAgE used it more compare to FMS or HaSS. The proportion of projects across faculty that used home storage is the same.

Table 3.5b: Cross tabulation of faculty with data storage location

			Data Storage Location								
			ISS managed systems	Academic Unit managed systems	Project managed systems	Personal systems	Other (please specify)	Off-campus managed system	External Cloud (IaaS) e.g. Amazon/EC2	External Service (SaaS) e.g. SurveyMonkey	At home
Faculty	HaSS	Count	14	4	5	10	3	2	1	1	9
		Column N %	17.9	8.5	22.7	19.2	25.0	18.2	16.7	50.0	33.3
	FMS	Count	40	27	9	21	6	7	2	1	9
		Column N %	51.3	57.4	40.9	40.4	50.0	63.6	33.3	50.0	33.3
	SAgE	Count	24	16	8	21	3	2	3	0	9
		Column N %	30.8	34.0	36.4	40.4	25.0	18.2	50.0	.0	33.3

Chi-Square 27.44, df=18 p=0.071

3.6 Length of Keeping Data

- **How long do you intend to keep data for?**

The majority of projects (nearly 43%) intend to keep data for 5-10 years. Just over 29% intend to keep data for 10-25 years and just over 17% intend to keep data for more than 25 years. Eleven percent wants to keep data for 1-5 years. It is worth noting that no project intends to keep data for less than a year (figure 3.6).

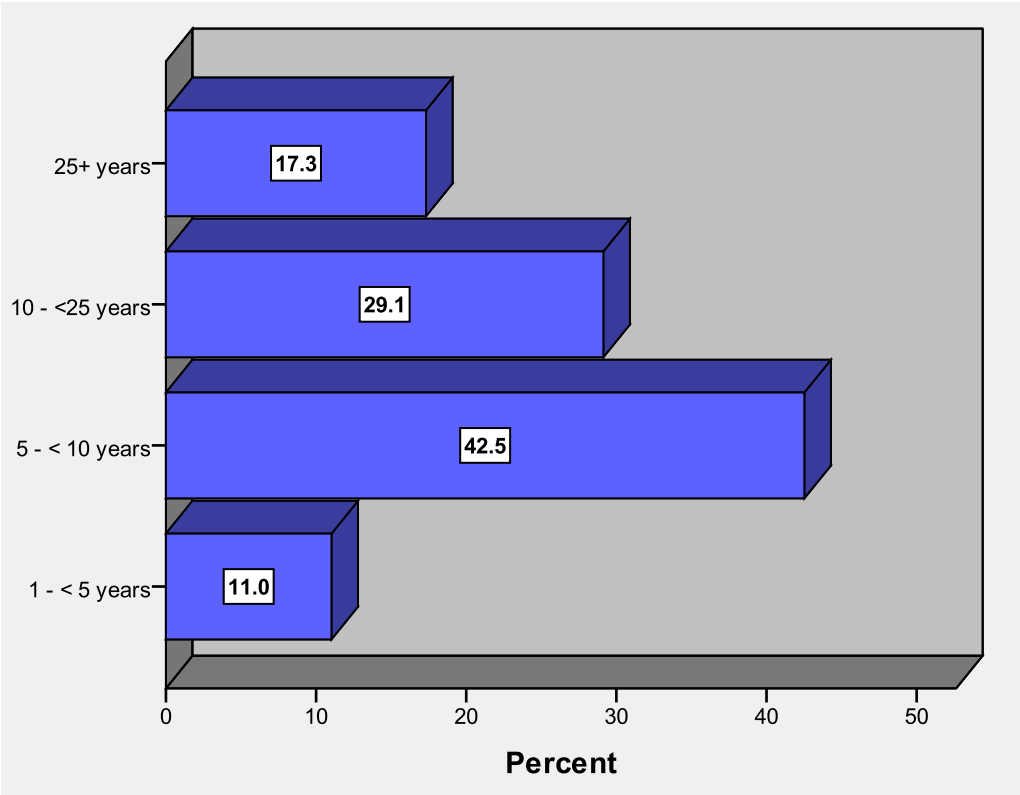

Figure 3.6: Length of keeping data

There is a statistically significant association between how long projects intend to keep data and faculty (Chi-square=13.88, df=6 and p-0.031 (< 0.05)). The proportions of projects from HaSS, FMS and SAgE who intend to keep data for between 5 - 10 years are 40.9%, 46.2% and 38.2% respectively. For a longer storage period that is 25+ the corresponding percent are 36.4%, 10.8% and 17.6% for HaSS, FMS and SAgE respectively. A higher proportion of projects in HaSS want to keep data for longer compared to projects from the other two faculties. See table 3.6 for details.

Table 3.6: Cross tabulation of faculty with length of keeping data

			How long do you intend to keep the data for?				
			1 - < 5 years	5 - < 10 years	10 - <25 years	25+ years	Total
Faculty	HaSS	Count	4	9	1	8	22
		% within Faculty	18.2%	40.9%	4.5%	36.4%	100.0%
	FMS	Count	5	30	23	7	65
		% within Faculty	7.7%	46.2%	35.4%	10.8%	100.0%
	SAgE	Count	4	13	11	6	34
		% within Faculty	11.8%	38.2%	32.4%	17.6%	100.0%
Total		Count	13	52	35	21	121
		% within Faculty	10.7%	43.0%	28.9%	17.4%	100.0%

Chi-Square 13.88, df=6 p=0.031

3.7 Additional Details on Data Section

Projects were given the opportunity to specify details whenever they select *Other*. Projects who do not use any of the data storage location mentioned above stores their data onsite (external hard drives in offices, USB drives, etc), offsite (PC at home, portable drives, NHS computers, etc.). Some projects are migrating data to collaborative platforms supported by the Digital Institute.

4 Research Data Management

4.1 Multiple Copies of Data

- **Are there multiple copies of the data (i.e. backed up)?**

The majority of projects (67.2%) have multiple copies of data backed up while nearly 26% have only partial back up. A quarter of projects have only partial back up of research data; nearly 5% don't have multiple backed up of their data and 2.3% of projects don't know whether they have multiple backed up of data (table 4.1a).

Table 4.1a: Multiple copies of backup data

Multiple backed up data		Frequency	Percent
Valid	I don't know	3	2.3
	No	6	4.7
	Partial back up	33	25.8
	Yes	86	67.2
	Total	128	100.0

Data backed practice of projects by faculty is shown on table 4.1b below. There is no statistically significant association between data backed up practice and faculty with chi-square=6.30, df=6, p=0.394 (>0.05). The percent of projects with multiple backed up copies of data are 63.6%, 66.7% and 70.6% for HaSS, FMS and SAgE respectively. The proportion of projects with partial backed up of data is very similar for HaSS and SAgE (31.8% vs. 29.4%); FMS is a bit lower at 21.2%.

Table 4.1b: Cross tabulation of faculty and data backup

			Are there multiple copies of the data (i.e. backed up)?				
			I don't know	No	Partial back up	Yes	Total
Faculty	HaSS	Count	0	1	7	14	22
		% within Faculty	.0%	4.5%	31.8%	63.6%	100.0%
	FMS	Count	3	5	14	44	66
		% within Faculty	4.5%	7.6%	21.2%	66.7%	100.0%
	SAgE	Count	0	0	10	24	34
		% within Faculty	.0%	.0%	29.4%	70.6%	100.0%
Total		Count	3	6	31	82	122
		% within Faculty	2.5%	4.9%	25.4%	67.2%	100.0%

Chi-Square=6.30, df=6 p=0.394

4.2 Testing Retrieving backed up data and sharing data

- **Have you successfully tested retrieving data from backup?**
- **Do you share or potentially share data with other in the University?**
- **Do you share or potentially share data with people external to the University?**

Nearly 49% of the projects have not tested how successful it will be to retrieve backed up data, while nearly 46% have tested it and about 6% said they don't know. Just over 73% of projects share their data with others in the University while nearly 24% don't share data and about 3% don't know. For data sharing external to the University 47.2% of projects said they don't while and 50.4% said they do and over 2% said they don't know. It is good to know that a high proportion of projects within University share research data. See table 4.2a.

Table 4.2a: Testing retrieving backup data and sharing data

Questions		I don't know	No	Yes
Have you successfully tested retrieving data from the backup?	Frequency	7	57	61
	Precent	5.6	45.6	48.8
Do you share or potentially share data with others in the University?	Frequency	4	29	90
	Precent	3.3	23.6	73.2
Do you share or potentially share data with people external to the University?	Frequency	3	60	64
	Precent	2.4	47.2	50.4

Now let us look at each of these questions by faculty.

The proportion of projects that have successfully tested retrieving backed up data is shown on the table below. The percentages are 61.9%, 47.7%, and 39.4% for HaSS, FMS and SAgE respectively (table 4.2b). However there is no statistically significant association between retrieval of backed up data with faculty (Chi-Square=3.93, df=4 p=0.419 (>0.05)).

Table 4.2b: Cross tabulation of faculty with retrieving backup data

			Have you successfully tested retrieving data from the backup?			
			I don't know	No	Yes	Total
Faculty	HaSS	Count	1	7	13	21
		% within Faculty	4.8%	33.3%	61.9%	100.0%
	FMS	Count	5	29	31	65
		% within Faculty	7.7%	44.6%	47.7%	100.0%
	SAgE	Count	1	19	13	33
		% within Faculty	3.0%	57.6%	39.4%	100.0%
Total		Count	7	55	57	119
		% within Faculty	5.9%	46.2%	47.9%	100.0%

Chi-Square=3.93, df=4 p=0.419

A higher proportion (91.2%) of projects from SAgE share data with others in the University; the corresponding figures for HaSS and FMS are 60.0% and 65.1% respectively (table 4.2c). There is a statistically significant association between data sharing with others within the University by faculty (Chi-Square=17.54, df=4 p=0.002 (<0.05)). Data sharing is something to be desired and projects in SAgE seem to have good practice on this. Details are shown on the table below.

Table 4.2c: Cross tabulation of faculty with data sharing within the University

			Do you share or potentially share data with others in the University?			
			I don't know	No	Yes	Total
Faculty	HaSS	Count	3	5	12	20
		% within Faculty	15.0%	25.0%	60.0%	100.0%
	FMS	Count	1	21	41	63
		% within Faculty	1.6%	33.3%	65.1%	100.0%
	SAgE	Count	0	3	31	34
		% within Faculty	.0%	8.8%	91.2%	100.0%
Total		Count	4	29	84	117
		% within Faculty	3.4%	24.8%	71.8%	100.0%

Chi-Square=17.54, df=4 p=0.002

Just like data sharing within the University, there is a statistically significant association between data sharing external to the University by faculty (Chi-Square=20.65, df=4 p=0.001 (<0.05)). The proportion of projects in SAgE that do this is 70.6%; the figures for HaSS and FMS are 57.1% and 37.9% respectively. Details are shown on table 4.2d below.

Table 4.2d: Cross tabulation of faculty with data sharing external to the University

			Do you share or potentially share data with people external to the University?			
			I don't know	No	Yes	Total
Faculty	HaSS	Count	2	7	12	21
		% within Faculty	9.5%	33.3%	57.1%	100.0%
	FMS	Count	0	41	25	66
		% within Faculty	.0%	62.1%	37.9%	100.0%
	SAgE	Count	0	10	24	34
		% within Faculty	.0%	29.4%	70.6%	100.0%
Total		Count	2	58	61	121
		% within Faculty	1.7%	47.9%	50.4%	100.0%

Chi-Square=20.65, df=4 p=0.001

4.3 Data Sharing Agreement

- **If yes to the above, is there an agreement in place to govern the sharing?**

Of the 64 projects (50.4%) who share data external to the University the data sharing agreement in use can be found on table 4.3. The table show that nearly 30% don't have any agreement in place while nearly 33% have *other* types of agreement.

Table 4.3: Data sharing agreement

Data Sharing Agreement	Frequency	Percent
Codified	9	14.1
I Don't Know	4	6.3
No agreement	19	29.7
Non codified	10	15.6
Not applicable (no sharing)	1	1.6
Other agreement	21	32.8
Total	64	100.0

4.4 Research Data Management Plan

- **Do you have a research data management plan?**

About 23% of projects have a research data management plan while just over 39% do not. About 6% of projects don't know whether they have a plan and nearly 33% have a partial /informal plan. See the pie chart below (figure 4.4).

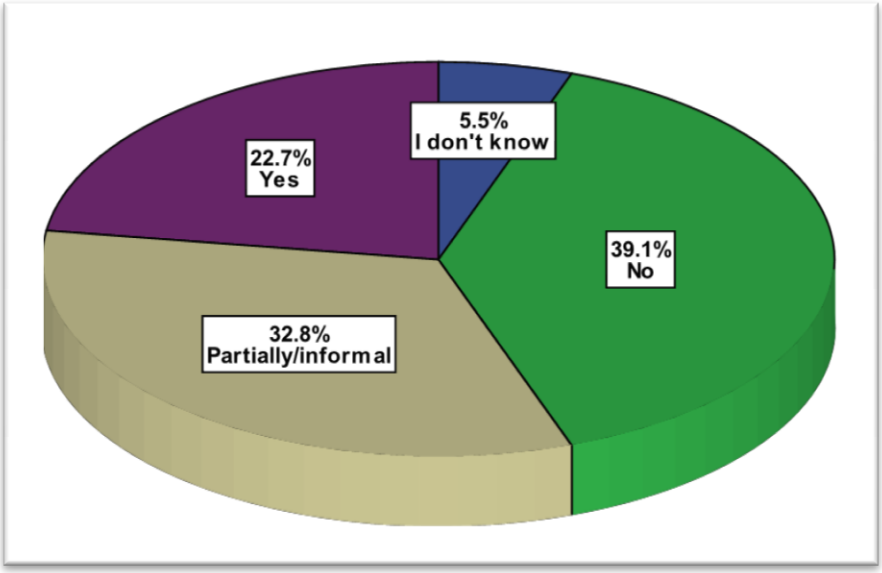

Figure 4.4: Research data management plan

The distribution of RDM plan by faculty is shown on the table below. There is no statistically significant association between RDM plans with faculty (Chi-Square=7.10, df=6 p=0.314 (>0.05)). However, a higher percent (27.3%) of projects from FMS have an RDM plan compared to only 11.8% and 18.3% for SAgE and HaSS respectively (table 4.4).

Table 4.4: Cross tabulation of faculty and research data management plan

			Do you have a research data management plan?				
			I don't know	No	Partially/informal	Yes	Total
Faculty	HaSS	Count	0	11	7	4	22
		% within Faculty	.0%	50.0%	31.8%	18.2%	100.0%
	FMS	Count	5	22	21	18	66
		% within Faculty	7.6%	33.3%	31.8%	27.3%	100.0%
	SAgE	Count	1	17	12	4	34
		% within Faculty	2.9%	50.0%	35.3%	11.8%	100.0%
Total		Count	6	50	40	26	122
		% within Faculty	4.9%	41.0%	32.8%	21.3%	100.0%

Chi-Square=7.10, df=6 p=0.314

4.5 Research Data Management Tools

- **Which tools do you use to manage your research data?**

As for data storage location, research data management tools was a multiply response question where projects were allowed to select all that apply to them. The total number of selection made across all the available research data management tools was 137. Note that 128 projects completed the online survey and on this question we had valid responses from 116 projects. Out of the 137 selections made 75 of them (54.7%) said they don't have any specific tool; 21.2% of projects used data file browser/search, just over 6% used repository file transfer and just over 7% used other systems that are specified in the next section. A small number of projects (1.5%) used data management planner and a similar proportion used metadata explorer. It is worth noting that for the projects that completed the online survey not a single project used Ontology or MyExperiment tools. See details on table 4.5a.

Table 4.5a: Research data management tools

RDM Tools	Responses		Percent of Cases
	N	Percent	
Data management planner	2	1.5	1.7
Metadata explorer	2	1.5	1.7
Data file browser/search	29	21.2	25.0
Data manager application	6	4.4	5.2
Repository file transfer	9	6.6	7.8
Other (please specify)	10	7.3	8.6
No specific tool	75	54.7	64.7
I don't know	4	2.9	3.4
Total	137	100.0	118.1

The distribution of RDM tools used by faculty is shown on the table below. From FMS 70.5% of projects do not have any specific RDM tools. The corresponding figures for HaSS and SAgE are 66.7% and 51.6% respectively (table 4.5b). However there is no statistically significant association between used of RMD tools and faculty (Chi-Square=19.50, df=16 p=0.243 (>0.05)). Nearly 39% of projects from SAgE used data file browser / search; the figures for HaSS and FMS are 19% and 21.3% respectively.

Table 4.5b: Cross tabulation of faculty and research data management tool

RDM Management Tools	Frequency / Percent	Faculty		
		HaSS	FMS	SAgE
Data management planner	Frequency	1	1	0
	Percent	4.8	1.6	.0
Metadata explorer	Frequency	1	0	1
	Percent	4.8	.0	3.2
Data file browser/search	Frequency	4	13	12
	Percent	19.0	21.3	38.7
Data manager application	Frequency	2	2	2
	Percent	9.5	3.3	6.5
Repository file transfer	Frequency	1	3	5
	Percent	4.8	4.9	16.1
Other (please specify)	Frequency	1	4	4
	Percent	4.8	6.6	12.9
No specific tool	Frequency	14	43	16
	Percent	66.7	70.5	51.6
I don't know	Frequency	1	1	2
	Percent	4.8	1.6	6.5
Total	Frequency	21	61	31

Chi-Square=19.50, df=16 p=0.243

4.6 Additional Tool Details

The projects that said they used other tools, the tools that they use include: besposke database, CARMEN (web based data management), dropbox, and Unix tools.

4.7 Deletion Policy

- **Do you have a deletion policy?**

The majority (46%) of projects said that they do not have any deletion policy this could mean that their policy is not to delete data. About 34% of projects said they don't delete data. It

seems that *I don't delete data* and *No* are the same thing. Just fewer than 17% of projects said they have a deletion policy and 3.2% said they don't know (table 4.7a).

Table 4.7a: Deletion policy

Deletion Policy		Frequency	Percent
Valid	I don't delete data	43	34.1
	I don't know	4	3.2
	No	58	46.0
	Yes	21	16.7
	Total	126	100.0
Missing	Missing	2	
Total		128	

The relationship between deletion policy by faculty is shown below. There is a statistically significant association between faculty and deletion policy (Chi-square=13.56, df=6 p=0.036 (>0.05)). A higher proportion of projects (61.8%) from SAgE do not have a deletion policy, the corresponding figures in HaSS and FMS are 42.9% and 36.9% respectively. Nearly a quarter of projects from FMS have a deletion possible, while for HaSS and SAgE the figures are only 14.3% and 2.9% respectively (table 4.7b).

Table 4.7b: Cross tabulation of faculty and deletion policy

			Do you have a deletion policy?				
			I don't delete data	I don't know	No	Yes	Total
Faculty	HaSS	Count	9	0	9	3	21
		% within Faculty	42.9%	.0%	42.9%	14.3%	100.0%
	FMS	Count	21	4	24	16	65
		% within Faculty	32.3%	6.2%	36.9%	24.6%	100.0%
	SAgE	Count	12	0	21	1	34
		% within Faculty	35.3%	.0%	61.8%	2.9%	100.0%
Total		Count	42	4	54	20	120
		% within Faculty	35.0%	3.3%	45.0%	16.7%	100.0%

Chi-Square=13.56, df=6 p=0.036

4.8 Need for Secure Data

- **How securely does your data need to be stored?**

Forty eight percent of the projects said their data have to be quite secure while just over a 27% of projects said their data have to be very secure and 24.2% said their data don't have to be secured. Percentages are shown on the pie chart below (figure 4.8).

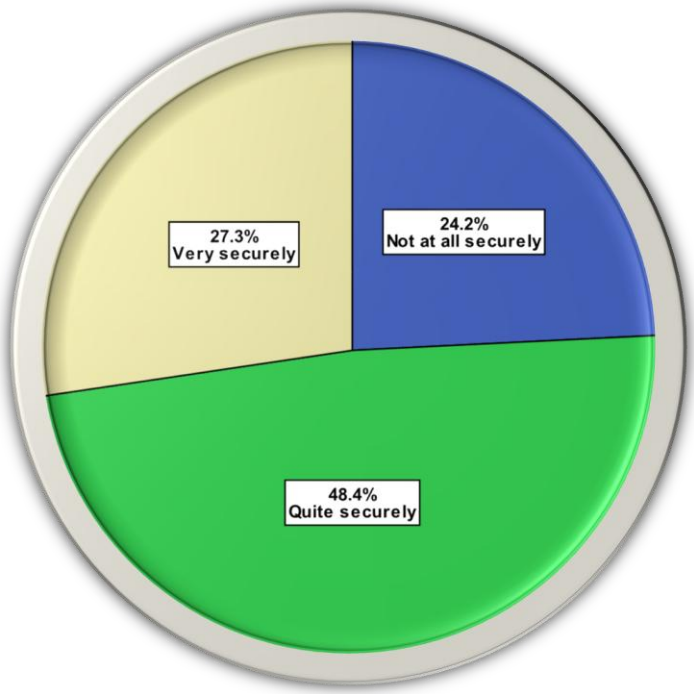

Figure 4.8: Need for secure data

Distribution of data security by faculty is shown on the table below. There is a statistically significant association between data security and faculty with a Chi-square value of 31.32, df of 4 and p=0.001 (<0.05). Nearly 56% of projects in SAgE said their data do not have to be stored securely, the figures for HaSS and FMS are only 18.2% and 10.6% respectively. Nearly 64% of projects from HaSS said their data have to be quite secured while the percent from FMS and SAgE are 47% and 35.3% respectively. About 42% of projects from FMS said that their data have to be very secured compared to the 18.2% and the 8.8% of projects from HaSS and SAgE respectively who said the same (table 4.8).

Table 4.8: Cross tabulation of faculty and need for secure data

			How securely does your data need to be stored?			
			Not at all securely	Quite securely	Very securely	Total
Faculty	HaSS	Count	4	14	4	22
		% within Faculty	18.2%	63.6%	18.2%	100.0%
	FMS	Count	7	31	28	66
		% within Faculty	10.6%	47.0%	42.4%	100.0%
	SAgE	Count	19	12	3	34
		% within Faculty	55.9%	35.3%	8.8%	100.0%
Total		Count	30	57	35	122
		% within Faculty	24.6%	46.7%	28.7%	100.0%

Chi-Square=31.32, df=4 p=0.001

4.9 Actual Security of Research Data

- **How securely do you actually store your data?**

The pattern for actual data security is very similar to the need of security discussed in the preceding section. As shown on the table below about 58% of projects said they store their data quite securely, over 25% said very securely and 16% said not at all securely (table 4.9a).

Table 4.9a: Actual security of research data

Actual Data Security		Frequency	Percent
Valid	Not at all securely	20	16.0
	Quite securely	73	58.4
	Very securely	32	25.6
	Total	125	100.0
Missing	Missing	3	
Total		128	

The distribution of actual data security in practice by faculty is shown on the data below. There is a statistically significant association between actual data security by faculty with a Chi-Square value of 17.66, df of 4 and p=0.001 (<0.05). Nearly 76% of projects in SAgE said their data is quite secure, next is HaSS with 65% and then FMS with 47.7%. FMS projects (40%) said their data is very secure, followed HaSS with 15% and then SAgE with only 2.9% (table 4.9b).

Table 4.9b: Cross tabulation of faculty and actual security of research data

			How securely do you actually store your data?			
			Not at all securely	Quite securely	Very securely	Total
Faculty	HaSS	Count	4	13	3	20
		% within Faculty	20.0%	65.0%	15.0%	100.0%
	FMS	Count	8	31	26	65
		% within Faculty	12.3%	47.7%	40.0%	100.0%
	SAgE	Count	7	26	1	34
		% within Faculty	20.6%	76.5%	2.9%	100.0%
Total		Count	19	70	30	119
		% within Faculty	16.0%	58.8%	25.2%	100.0%

Chi-Square=17.66, df=4 p=0.001

4.10 Data Security Strategies

- **What strategies do you use to secure your data?**

Again as for data storage location and research data management tools this question was a multiply response question where projects were allowed to select all that apply to them. The total number of selection made across all the available data security strategies was 192. Note that 128 projects completed the online survey and on this question we had valid responses from 112 projects. Out of the 192 selections made 89 of them (46.4%) said they used password protection; 22.9% of projects used physical measures, 21.4% used anonymisation and 6.8% used encryption (table 4.10a). A small number of projects (2.6%) used other methods (see Data Management Section - additional details below).

Table 4.10a: Data security strategies

Strategies for Data Security	Responses		Percent of Cases
	Frequency	Percent	
Encryption	13	6.8	11.2
Anonymisation	41	21.4	35.3
Password protection	89	46.4	76.7
Physical measures	44	22.9	37.9
Other (please specify)	5	2.6	4.3
Total	192	100.0	165.5

The distribution of data security strategies by faculty is shown on the table below. There is a statistically significant association between data security strategies and faculty (Chi-Square=38.19, df=10, p=0.001 (<0.05)). Take encryption for example, 91.7% of the selection came from projects in FMS, 8.3% from SAgE and none from HaSS, see table 4.10b below. This pattern is repeated for the remaining data security strategies except for *Other* where a higher proportion of projects from HaSS used it compare to the other two faculties.

Table 4.10b: Cross tabulation of faculty and data security strategies

Faculty	Frequency / Percent	Data Security Strategies				
		Encryption	Anonymisation	Password protection	Physical measures	Other (please specify)
HaSS	Frequency	0	7	13	6	2
	Precent	.0	17.5	15.7	14.0	66.7
FMS	Frequency	11	33	49	25	1
	Precent	91.7	82.5	59.0	58.1	33.3
SAgE	Frequency	1	0	21	12	0
	Precent	8.3	.0	25.3	27.9	.0

Chi-Square=38.19, df=10 p=0.001

4.11 Policies and Legislation

- **Which guidance, policies and legislation are you aware of that covers your data?**

This was also a multiply response question where projects were ask to select all that apply. The total number of selection made across all the available policies and legislation was 355. Note that 128 projects completed the online survey and on this question we had valid responses from 109 projects. About 20% and 12% of projects are aware of DPA and FOI respectively. Nearly 12% of projects are aware of ISS Guidelines on Data Security. Just over 9% of projects are aware of Research Councils, Publication journals and NHS ethics/LREC policies and legislation (table 4.11a). About 2% of projects are aware of other policies and legislation (see Data Management Section - additional details below).

Table 4.11a: Policies and legislation

Policies and Legislation	Responses		Percent of Cases
	Frequency	Percent	
Data Protection Act	69	19.4	63.3
Freedom of Information	44	12.4	40.4
ISS Guidelines on Data Security	41	11.5	37.6
NHS requirements (please specify)	26	7.3	23.9
Research Councils	33	9.3	30.3
Publication journals	34	9.6	31.2
Local ethics/LREC	26	7.3	23.9
Other (please specify)	8	2.3	7.3
NHS ethics/NRES	33	9.3	30.3
University requirements (please specify)	26	7.3	23.9
Academic Unit requirements (please specify)	8	2.3	7.3
Reseach team requirements (please specify)	7	2.0	6.4
Total	355	100.0	325.7

The selection made on each policies and legislation according to faculty is shown on the table below. There is a statistically significant association between policies / legislation according to faculty with a Chi-Square value of 120.46, df of 24 and p=0.001 (<0.05). For example consider DPA, 68.2% of projects from FMS selected it while only 19.7% and 12.1% of projects from HaSS and SAgE respectively selected it. This is a very common pattern in the table except in *Other* where the figures are 57.1%, 28.6% and 14.3% for HaSS, FMS and SAgE respectively (table 4.11b).

Table 4.11b: Cross tabulation of faculty and policies / legislation

Policies and Legislation	Frequency / Row %	Faculty		
		HaSS	FMS	SAgE
Data Protection Act	Frequency	13	45	8
	Row %	19.7	68.2	12.1
Freedom of Information	Frequency	7	29	8
	Row %	15.9	65.9	18.2
ISS Guidelines on Data Security	Frequency	7	26	8
	Row %	17.1	63.4	19.5
NHS requirements (please specify)	Frequency	0	24	0
	Row %	.0	100.0	.0
Research Councils	Frequency	5	17	10
	Row %	15.6	53.1	31.3
Publication journals	Frequency	4	15	12
	Row %	12.9	48.4	38.7
Local ethics/LREC	Frequency	2	23	0
	Row %	8.0	92.0	.0
Other (please specify)	Frequency	4	2	1
	Row %	57.1	28.6	14.3
NHS ethics/NRES	Frequency	0	32	0
	Row %	.0	100.0	.0
University requirements (please specify)	Frequency	3	20	2
	Row %	12.0	80.0	8.0
Academic Unit requirements (please specify)	Frequency	1	7	0
	Row %	12.5	87.5	.0
Reseach team requirements (please specify)	Frequency	1	4	1
	Row %	16.7	66.7	16.7

Chi-Square=120.46, df=24 p=0.001

4.12 Policies and Legislation that would be useful to Projects

- **Which guidance, policies and legislation would it be most useful for the University to provide you with further information on?**

The projects said it will be most useful for the University to provide further information on following policies and legislations:

- Data Protection Act (DPA)
- Freedom of Information Act (FOIA)
- ISS Guidelines on data security
- Funders minimum requirements
- NHS requirements
- University requirements
- Data management tools

4.13 Primary Responsibility for Research Data Management Support

- **Who has primary responsibility for research data management support?**

The majority of projects (nearly 64%) said that the PI should have the primary responsibility for research data management support. Next in line was the Research Associate with nearly 11% of projects saying they should have the responsibility. Only 0.8% of projects think that PhD students should have the primary responsibility for research data management support. Nearly 2% of projects think that research administrators should be in charge and another 2% think that technical support staff should be in charge. Almost 7% of projects said other officers should have the primary responsibility (see Data Management Section - additional details below). See table 4.13a for details.

Table 4.13a: Primary responsibility for RDM support

Current Reponsible Officer		Frequency	Percent
Valid	Computing Support Officer	12	10.1
	PhD student	1	.8
	Principal Investigator	76	63.9
	Research Administrator	2	1.7
	Research Assistant	5	4.2
	Research Associate	13	10.9
	Technical support	2	1.7
	Other (please specify)	8	6.7
	Total	119	100.0
Missing	Missing	9	
Total		128	

The views of projects according to faculty are shown on table 4.13b below. Across faculty the majority of projects said that the PI should have the primary responsibility, the figures are 66.7%, 69.8% and 50% for HaSS, FMS and SAgE respectively. However, there is no statistically significant association between primary responsibility and faculty with Chi-Square value of 21.32, df of 14 and p=0.073 (>0.05).

Table 4.13b: Cross tabulation of faculty and primary responsibility for RDM support

			Who has primary responsibililty for research data management support?								
			Computing Support Officer	Other (please specify)	PhD student	Principal Investigator	Research Administrator	Research Assistant	Research Associate	Technical support	Total
Faculty	HaSS	Count	0	3	0	14	0	1	2	1	21
		% within Faculty	.0%	14.3%	.0%	66.7%	.0%	4.8%	9.5%	4.8%	100.0%
	FMS	Count	10	2	0	44	0	1	5	1	63
		% within Faculty	15.9%	3.2%	.0%	69.8%	.0%	1.6%	7.9%	1.6%	100.0%
	SAgE	Count	2	3	1	16	1	3	6	0	32
		% within Faculty	6.3%	9.4%	3.1%	50.0%	3.1%	9.4%	18.8%	.0%	100.0%
Total		Count	12	8	1	74	1	5	13	2	116
		% within Faculty	10.3%	6.9%	.9%	63.8%	.9%	4.3%	11.2%	1.7%	100.0%

Chi-Square=21.32 df=14 p=0.073

4.14 Future Responsibility for Research Data Management Support

- **In future, who should have primary responsibility for research data management support?**

As for the last question, the majority of projects (nearly 55%) believe that going forward the PI should still have the primary responsibility for research data management support. All the other officers got less than 10% of the vote as can be found on table 4.14a below except for computing support officer where just over 17% of projects think they should have the primary responsibility for RDM support.

Table 4.14a: Future responsibility for RDM support

Future Reponsible Officer		Frequency	Percent
Valid	Computing Support Officer	20	17.1
	PhD Student	1	.9
	Principal Investigator	64	54.7
	Research Administrator	4	3.4
	Research Assistant	3	2.6
	Research Associate	10	8.5
	School Manager	3	2.6
	Technical support	4	3.4
	Other (please specify)	8	6.8
	Total	117	100.0
Missing	Missing	11	
Total		128	

The pattern according to faculty is very similar to the last question where 52.6%, 62.9% and 42.4% of projects from HaSS, FMS and SAgE respectively think that the PI should have the primary responsibility in the future for RDM support as shown on table 4.14b below.

However, there is no statistically significant association between primary responsibility and faculty with Chi-Square value of 22.23, df of 16 and p=0.136 (>0.05).

Table 4.14b: Cross tabulation of faculty and future responsibility for RDM support

			In the future, who should have primary responsibility for research data management support?									
			Computing Support Officer	Other (please specify)	PhD Student	Principal Investigator	Research Administrator	Research Assistant	Research Associate	School Manager	Technical support	Total
Faculty	HaSS	Count	1	3	0	10	0	0	3	0	2	19
		% within Faculty	5.3%	15.8%	.0%	52.6%	.0%	.0%	15.8%	.0%	10.5%	100.0%
	FMS	Count	13	2	0	39	0	1	4	2	1	62
		% within Faculty	21.0%	3.2%	.0%	62.9%	.0%	1.6%	6.5%	3.2%	1.6%	100.0%
	SAgE	Count	6	3	1	14	2	2	3	1	1	33
		% within Faculty	18.2%	9.1%	3.0%	42.4%	6.1%	6.1%	9.1%	3.0%	3.0%	100.0%
Total		Count	20	8	1	63	2	3	10	3	4	114
		% within Faculty	17.5%	7.0%	.9%	55.3%	1.8%	2.6%	8.8%	2.6%	3.5%	100.0%

Chi-Square=22.23 df=16 p=0.136

4.15 Research Data Management Training Session / Materials Awareness

- **What research data management (or closely related) training sessions or training materials are you aware of?**

It is worth noting that 77 projects making 60.2% of the responses did not provide an answer to this question, this could mean that the projects are not aware of any RDM training sessions / materials. Only 51 projects (39.8%) answered this question. Of those that answered the question 43 of them making about 84% are not aware of any research data management training sessions or materials that exist. Only 5 projects (1 from HaSS, 3 from FMS and 1 from SAgE) said they are aware of training sessions and materials on RDM (table 4.15).

Table 4.15: Awareness of RDM session / materials

Awareness of RDM session / materials	Frequency	Percent
Yes	5	3.9
No	46	35.9
Missing (no answer)	77	60.2
Total	128	

4.16 Data Management Section - additional details

Projects that answered *other* on data security strategies, policies and legislation awareness and any of the proceeding sections provided additional details. The PI is in charge of data management. PI also has a role to play to ensure that data is secured even though this responsibility is delegated to members in the research team. Ethics approval is sought on most projects and data security depends on the project. For example patient data is very secured if non-anonymised using encryption or password; anonymised data only mildly

secured. Other type of data e.g. astronomical data is not confidential or personal so security is not an issue.

5 Data Repositories

5.1 Making Data Accessible and Discoverable at end of project

- **At the end of your project, are you happy to have your data publically discoverable and accessible?**

As can be seen on the figure 5.1 below the answers selected by projects for this question are almost evenly distributed. About one quarter of projects said they are partially happy to make their research data publically available at the end of the project. About 22% said 'no' while nearly 21% said 'yes'. Nearly 14% said maybe and 17.6% are willing to do this if participants are anonymised.

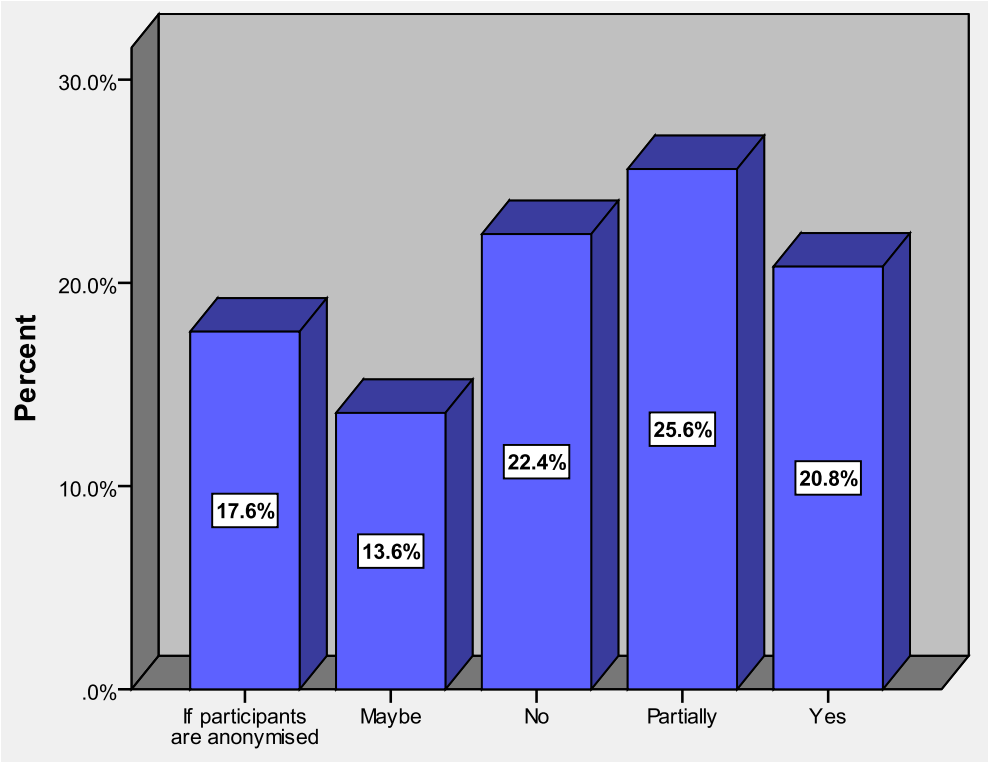

Figure 5.1: Making data accessible and discoverable at the end of project

The answers to this question by faculty are shown on the table 5.1 below. There is a statistically significant association between making data accessible and faculty with a Chi-Square value of 15.65, df of 8 and p=0.046 (<0.05). Indicating that there is strong evidence beyond chance making data accessible differs by faculty. For instance 28.6%, 14.1% and 26.5% of projects from HaSS, FMS and SAgE respectively said yes to make data accessible

at the end of the project. Notice the proportion for HaSS and SAgE almost double that for FMS. The figures for 'partially' are 19.0%, 28.1%, and 20.6% from HaSS, FMS and SAgE respectively.

Table 5.1: Cross tabulation of Faculty and Making data accessible

			At the end of your project, are you happy to have your data publically discoverable and accessible?					
			If participants are anonymised	Maybe	No	Partially	Yes	Total
Faculty	HaSS	Count	5	2	4	4	6	21
		% within Faculty	23.8%	9.5%	19.0%	19.0%	28.6%	100.0%
	FMS	Count	17	8	12	18	9	64
		% within Faculty	26.6%	12.5%	18.8%	28.1%	14.1%	100.0%
	SAgE	Count	0	7	11	7	9	34
		% within Faculty	.0%	20.6%	32.4%	20.6%	26.5%	100.0%
Total		Count	22	17	27	29	24	119
		% within Faculty	18.5%	14.3%	22.7%	24.4%	20.2%	100.0%

Chi-Square=15.65 df=8 p=0.046

5.2 Data Deposited in Repository

- **Have you ever deposited any of your data into a data repository?**

The distribution of answers according to the use of data repository by projects is shown on figure 5.2 below. The majority of projects (73%) have not deposited any of their data in a data repository. Nearly 13% have done so because it is a requirement for them and 12% chose to do so. A small proportion (2.4%) of projects doesn't know.

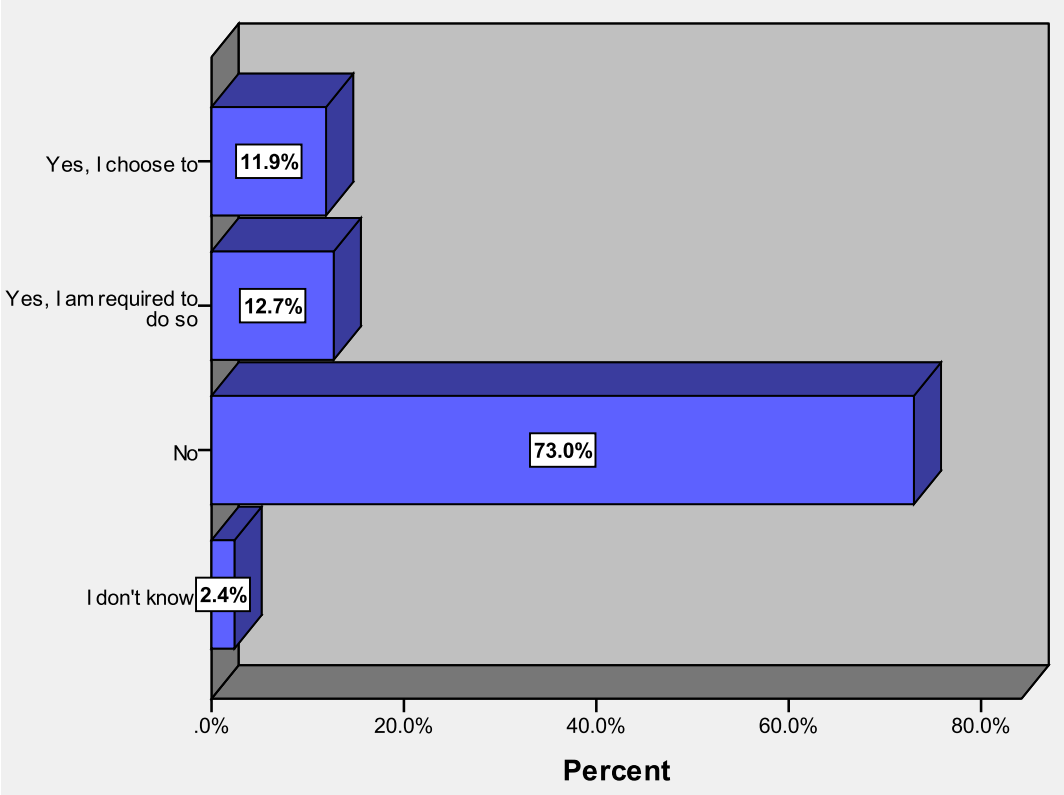

Figure 5.2: Data deposited in a data repository

The use of data repository according to faculty is shown on table 5.2 below. There is no statistically significant association between use of data repository and faculty (Chi-Square=7.81, df=6, p=0.248 (>0.05)). The percent of projects that have not used data repository are 52.4%, 76.9% and 73.5% from HaSS, FMS and SAgE respectively. The percent of projects that are required to do so are 23.8%, 7.7% and 17.6% from HaSS, FMS and SAgE respectively. It is more of a requirement for projects in HaSS compared to projects in FMS and SAgE.

Table 5.2: Cross tabulation of faculty with data repository

			Have you ever deposited any of your data into a data repository?				
			I don't know	No	Yes, I am required to do so	Yes, I choose to	Total
Faculty	HaSS	Count	1	11	5	4	21
		% within Faculty	4.8%	52.4%	23.8%	19.0%	100.0%
	FMS	Count	1	50	5	9	65
		% within Faculty	1.5%	76.9%	7.7%	13.8%	100.0%
	SAgE	Count	1	25	6	2	34
		% within Faculty	2.9%	73.5%	17.6%	5.9%	100.0%
Total		Count	3	86	16	15	120
		% within Faculty	2.5%	71.7%	13.3%	12.5%	100.0%

Chi-Square=7.81 df=6 p=0.248

5.3 Willing to Submit Data to a Data Repository

- **Would you be willing to submit your data to a data repository?**

As shown on figure 5.3 below one third of projects are willing to submit data to a data repository while about 14% are not willing to do so; about 15% don't know and about 14% are happy to do so if participants are anonymised. About a quarter of projects are partially willing to submit data to a data repository.

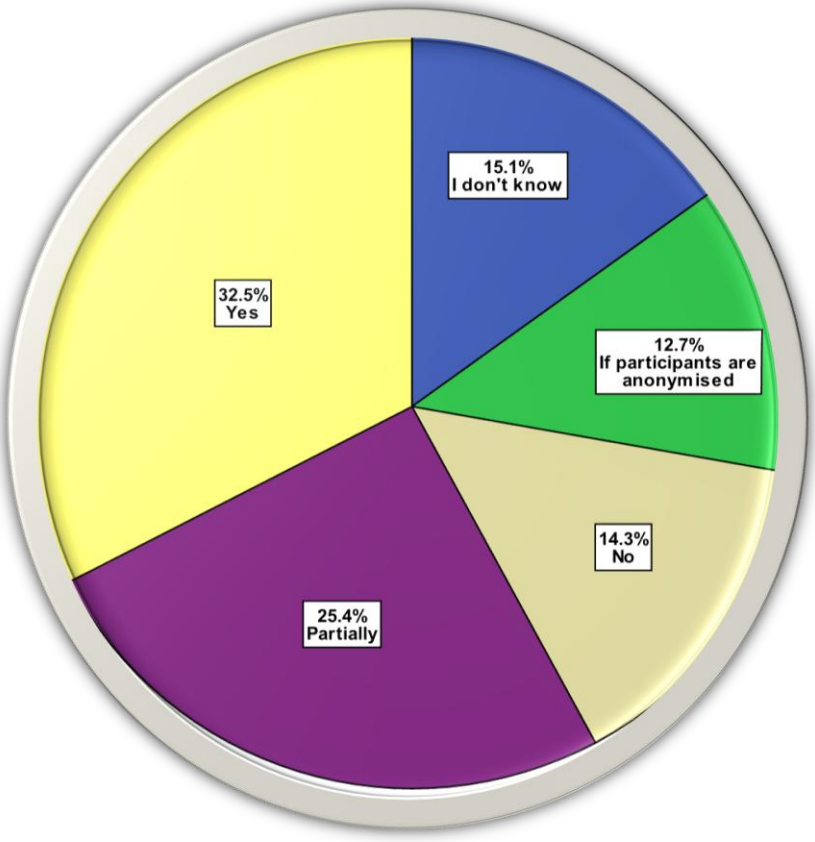

Figure 5.3: Willing to submit data to a data repository

As shown on table 5.3 below, of those who said they are willing to submit data into a data repository 33.3%, 24.6% and 44.1% were from HaSS, FMS and SAgE respectively. Even though the proportion from SAgE is bigger than the other faculties there is no statistically significant association between wiliness to submit data to data repository and faculty (Chi-Square=9.51, df=8 p=0.307 (>0.05)). For those who said no the corresponding percent across faculty are 14.3%, 13.8% and 17.6% from HaSS, FMS and SAgE respectively.

Table 5.3: Cross tabulation of faculty and willing to submit data to a data repository

			Would you be willing to submit your data to a data repository?					
			I don't know	If participants are anonymised	No	Partially	Yes	Total
Faculty	HaSS	Count	3	1	3	7	7	21
		% within Faculty	14.3%	4.8%	14.3%	33.3%	33.3%	100.0%
	FMS	Count	11	12	9	17	16	65
		% within Faculty	16.9%	18.5%	13.8%	26.2%	24.6%	100.0%
	SAgE	Count	5	1	6	7	15	34
		% within Faculty	14.7%	2.9%	17.6%	20.6%	44.1%	100.0%
Total		Count	19	14	18	31	38	120
		% within Faculty	15.8%	11.7%	15.0%	25.8%	31.7%	100.0%

Chi-Square=9.51 df=8 p=0.307

5.4 Stage in Data's Lifecycle to Submit Data to Repository

- **At what stage in the data's lifecycle would you submit data to the repository?**

An overwhelming majority of projects (nearly 80%) are happy to submit data to a repository at the publication stage while just over 15% said they wouldn't submit data at all. A small proportion (1.6%) said at the collection and processing stages (figure 5.4).

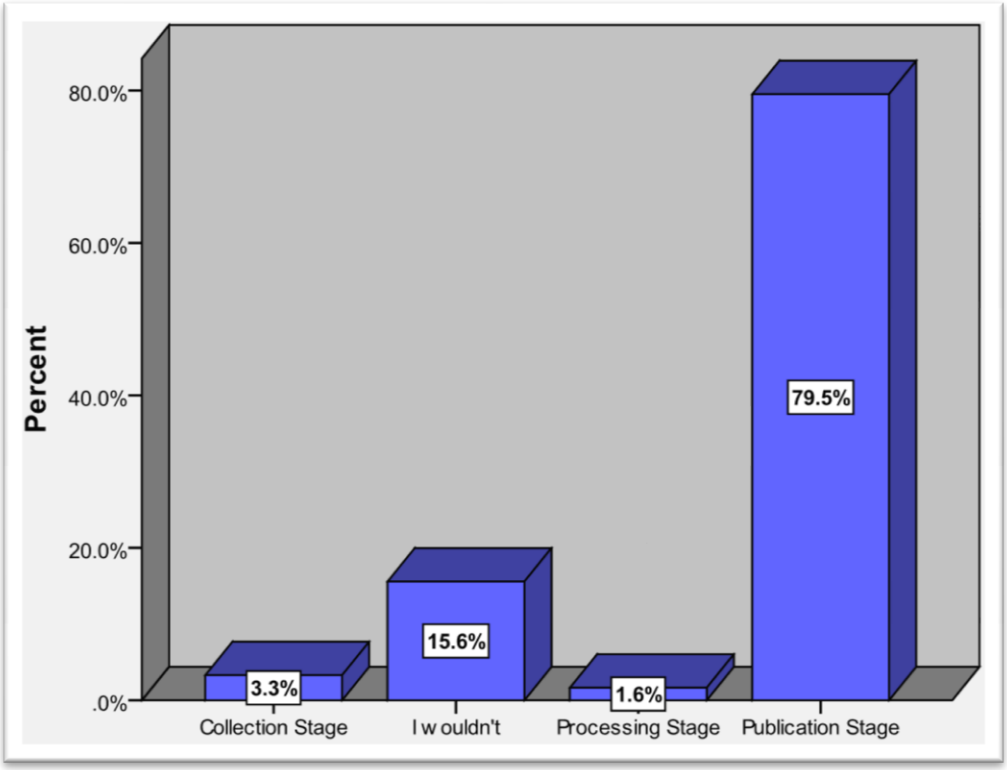

Figure 5.4: Data's lifecycle stage to submit data to the repository

The distribution by faculty of those who said they are willing to submit data at publication stage is shown on table 5.4 below. The proportions of projects by faculty are 76.2%, 82.3% and 75.8% from HaSS, FMS and SAgE respectively. There is however no statistically significant association between stage of data submission to a data repository and faculty.

Table 5.4: Stage in Data's Lifecycle to Submit Data to Repository

			At what stage in the data's lifecycle would you submit your data to the repository?				
			Collection Stage	I wouldn't	Processing Stage	Publication Stage	Total
Faculty	HaSS	Count	1	3	1	16	21
		% within Faculty	4.8%	14.3%	4.8%	76.2%	100.0%
	FMS	Count	1	10	0	51	62
		% within Faculty	1.6%	16.1%	.0%	82.3%	100.0%
	SAgE	Count	2	6	0	25	33
		% within Faculty	6.1%	18.2%	.0%	75.8%	100.0%
Total		Count	4	19	1	92	116
		% within Faculty	3.4%	16.4%	.9%	79.3%	100.0%

Chi-Square=6.16 df=6 p=0.425

5.5 Making Data Available after Publication

- **Thinking now, specifically about data at publication, how long after publication would you be willing to make the data supporting that publication available?**

The majority of projects (40.7%) said immediately, while about 19% said 2 – 5 years. About 7% said never and about 11% said within 1 month - 12 months. Nearly 9% said *other* and what they specify can be found in section 5.7 below. A very small proportion (1.6%) said after retirement (figure 5.5).

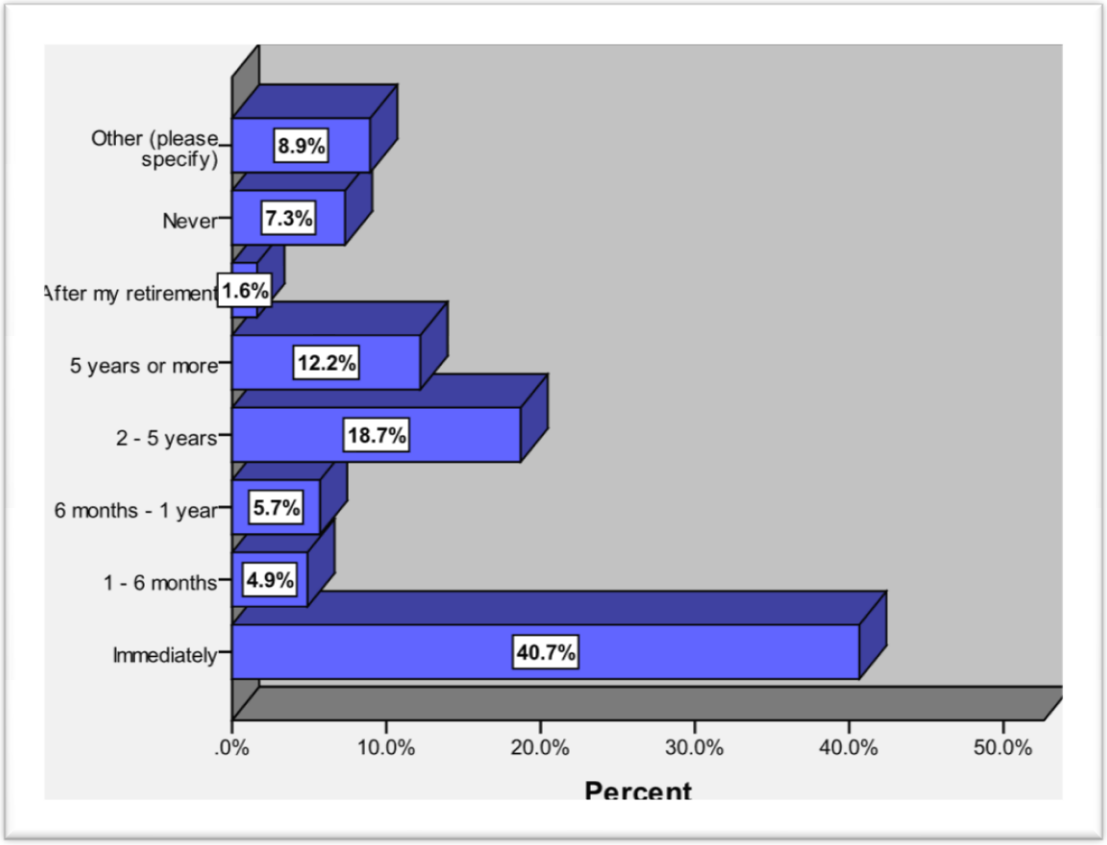

Figure 5.5: Making data available after publication

Of those who said 'immediately' 35%, 39.7% and 47.1% were from HaSS, FMS and SAgE respectively. For those who said '2-5 years' the corresponding percents are 25%, 23.8% and 5.9% respectively. However there is no statistically significant association between making data available after publication by faculty (Chi-Square=16.14, df=14 and p=0.305 (>0.05)).

5.6 Share Data if you control access to it

- **Would you be more likely to share the data if you controlled who could access it?**

As shown on the 3D pie chart in figure 5.6 below 41% of projects are willing to share data if they have control over who can access the data. Just over 27% said they still will not share the data and nearly 32% said maybe.

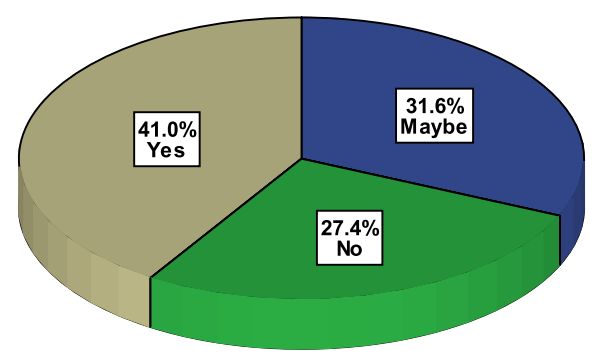

Figure 5.6: Data sharing if you control access to it

The distribution of answers by projects according to faculty is shown on table 5.6 below. Of those who said 'yes' 26.3%, 46.8%, 26.7% are from HaSS, FMS and SAgE respectively. The percent for HaSS and SAgE are similar while that for FMS is bigger, however, there is no statistically significant association between sharing data and faculty (Chi-Square=5.73, df=4, p=0.223 (>0.05)).

Table 5.6: Cross tabulation of faculty with data sharing if you control access to it

			Would you be more likely to share the data if you controlled who could access it?			
			Maybe	No	Yes	Total
Faculty	HaSS	Count	8	6	5	19
		% within Faculty	42.1%	31.6%	26.3%	100.0%
	FMS	Count	19	14	29	62
		% within Faculty	30.6%	22.6%	46.8%	100.0%
	SAgE	Count	10	12	8	30
		% within Faculty	33.3%	40.0%	26.7%	100.0%
Total		Count	37	32	42	111
		% within Faculty	33.3%	28.8%	37.8%	100.0%

Chi-Square=5.73 df=4 p=0..223

5.7 Data Repositories Section – Additional Details

Additional details for those who specify *other* in section 5.5. Generally the comment seems to indicate that projects are willing to make data available provided it is not commercially sensitive and participants have given their consent. Projects will share some data when it has been published / patented. However some data will never be shared such as data on national security or due to funder restrictions. Some projects think that data per se is meaningless without the accompanying metadata and producing this is time consuming and some data are only useful within the context of collection.

6 Intellectual Property

6.1 Intellectual Property Rights (IPR)

- **Who has the intellectual property rights for your research data?**

As shown on table 6.1 below, there is no clear consensus on intellectual property rights. Just over 30% of projects believe that it is owned by the University and about 17% of projects do not know; just over 19% think it belongs to their research group and 10.4% said *other*. For about 7% and 15% of projects it belonged to the funder and the researcher respectively.

Table 6.1: Intellectual property rights

Intellectual Property Rights		Frequency	Percent
Valid	Another group/organisation	1	.8
	I don't know	21	16.8
	Me	19	15.2
	My Funder	9	7.2
	My research group	24	19.2
	Other (please specify)	13	10.4
	The University	38	30.4
	Total	125	100.0
Missing	Missing	3	
Total		128	

Looking at the answers on IPR issue by faculty as shown on table 6.1b below of those who said it belong to the University 19%, 34.4%, and 26.5% are from HaSS, FMS and SAgE respectively. For those who said it belong to their research group 23.8%, 20.3%, and 11.8% are from HaSS, FMS and SAgE respectively. The corresponding figures for projects that think IPR belong to them 33.3%, 10.9% and 11.8% are from HaSS, FMS and SAgE respectively. However there is no statistically significant association between IPR and faculty with Chi-Square of 16.83, df=12 and p=0.156 (>0.05).

Table 6.1b: Cross tabulation of faculty and intellectual property rights

			Who has the intellectual property rights for your research data?							Total
			Another group/organisation	I don't know	Me	My Funder	My research group	Other (please specify)	The University	
Faculty	HaSS	Count	0	2	7	2	5	1	4	21
		% within Faculty	.0%	9.5%	33.3%	9.5%	23.8%	4.8%	19.0%	100.0%
	FMS	Count	0	10	7	6	13	6	22	64
		% within Faculty	.0%	15.6%	10.9%	9.4%	20.3%	9.4%	34.4%	100.0%
	SAgE	Count	1	9	4	1	4	6	9	34
		% within Faculty	2.9%	26.5%	11.8%	2.9%	11.8%	17.6%	26.5%	100.0%
Total		Count	1	21	18	9	22	13	35	119
		% within Faculty	.8%	17.6%	15.1%	7.6%	18.5%	10.9%	29.4%	100.0%

Chi-Square=16.83 df=12 p=0.156

6.2 Funder

- **Who funded this project?**

The majority of the research projects were funded by either charity or research council 35.5% and 31.4% respectively, that is, total of nearly 67%. All the other sources of funding as can be seen on table 6.2a below are less than 10% each.

Table 6.2a: Funders

Funder		Frequency	Percent
Valid	Charity	43	35.5
	European Commission	6	5.0
	Industry/public corporations	4	3.3
	NHS/Department of Health	7	5.8
	Research Council	38	31.4
	UK government department	6	5.0
	Unfunded	8	6.6
	Other (please specify)	9	7.4
	Total	121	100.0
Missing	Missing	7	
Total		128	

The distribution of funders by faculty is shown on table 6.2b below. There is a statistically significant association between funder and faculty (Chi-Square=29.69, df=14 and p=0.008 (<0.05)). For HaSS 23.8% and 38.1% of projects were funded by charity and research councils respectively. For FMS the figures are 50% v 20.3% and for SAgE 15.2% v 51.5% respectively.

Table 6.2b: Cross tabulation of faculty and funders

			Who funded this project?								Total
			Charity	European Commission	Industry/public corporations	NHS/Department of Health	Other (please specify)	Research Council	UK government department	Unfunded	
Faculty	HaSS	Count	5	2	1	0	2	8	2	1	21
		% within Faculty	23.8%	9.5%	4.8%	.0%	9.5%	38.1%	9.5%	4.8%	100.0%
	FMS	Count	32	0	2	7	3	13	3	4	64
		% within Faculty	50.0%	.0%	3.1%	10.9%	4.7%	20.3%	4.7%	6.3%	100.0%
	SAgE	Count	5	3	1	0	3	17	1	3	33
		% within Faculty	15.2%	9.1%	3.0%	.0%	9.1%	51.5%	3.0%	9.1%	100.0%
Total		Count	42	5	4	7	8	38	6	8	118
		% within Faculty	35.6%	4.2%	3.4%	5.9%	6.8%	32.2%	5.1%	6.8%	100.0%

Chi-Square=29.69 df=14 p=0.008

6.3 Intellectual Property Section – Additional Details

Those who specify *other* in section 6.1 and 6.2 provided additional. Generally it is clear that the issue of IPR is complex and depends on the project. The majority of projects believe that IP is belongs to them, the University and the funder. Other sources of funding include overseas government, and institutes.

7 Questions, Comments and Further Information

Projects were given the opportunity to raise questions, make comments or seek further information. Twenty eight projects (just over 21.8%) took advantage of this offer. Most of the comments were on data storage space even though iridium is more on research data management. Some projects need more data storage space to prevent projects from using their own storage (servers, hard drives, pen drives, PCs, etc.). Some PIs felt that members of their team did not have the skills for data management.

Some projects were concerned that policies will be laid down that is not appropriate to all research data, for instance, astronomical data or data generated from computer simulation.

Some projects could not understand or fully appreciate the importance / need for data management plans.