

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Metadata is a Love Note to the Future: UK Higher Education Research Data Management Survey

PLEASE CITE THE PUBLISHED VERSION

<http://dx.doi.org/10.6084/m9.figshare.816938>

VERSION

VoR (Version of Record)

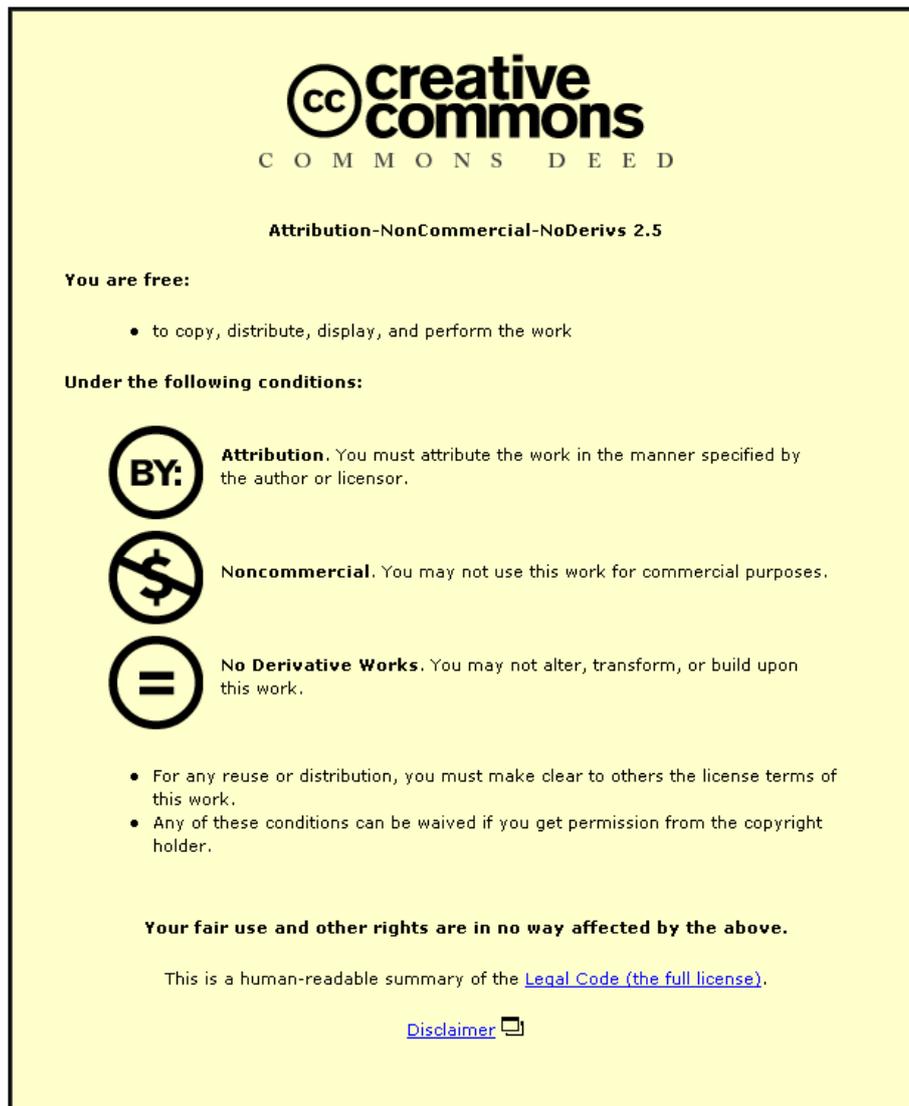
LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Hamilton, Martin, and Sue Manuel. 2019. "Metadata Is a Love Note to the Future: UK Higher Education Research Data Management Survey". figshare. <https://hdl.handle.net/2134/13526>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

BY: **Attribution.** You must attribute the work in the manner specified by the author or licensor.

Noncommercial. You may not use this work for commercial purposes.

No Derivative Works. You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

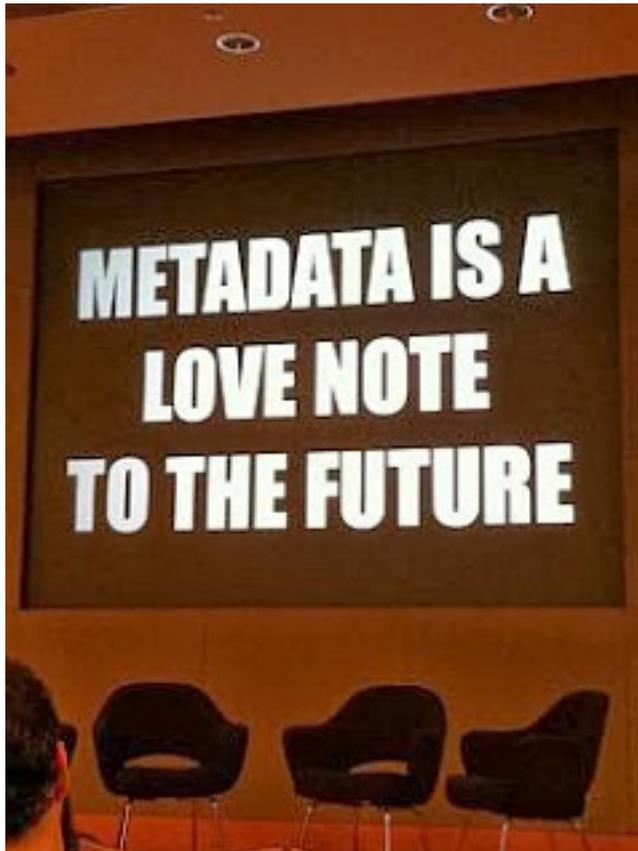
Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Metadata is a Love Note to the Future: UK Higher Education Research Data Management (RDM) Survey



Martin Hamilton, Sue Manuel
Loughborough University

October 2013

[Photo credit: Metadata is a Love Note to the Future, photo by [@kissane](#)]

1. Introduction	3
2. Executive Summary	4
3. Get The Data	6
4. Survey Responses	7
4.1 Does your institution have a Research Data Management policy?	7
4.2 Does your institute have a Research Data Service?	7
4.3 Scope of Research Data Service.....	7
4.4 Extent of data to be held in Research Data Service.....	8
4.5 Categories of information your Research Data Service stores.....	9
4.6 What software does your Research Data Service use?	9
4.7 How much storage is available for research data sets?	10
4.8 How much storage is each researcher allocated by default?	10
4.9 Does your institution charge for the Research Data Service?	11
4.10 Staffing associated with the Research Data Service.....	11
4.11 Support available for Research Data Management at your institution.....	13
5. Further Information	14
6. Survey Respondents	17

1. Introduction

Research Data Management is a hot topic for most UK HE institutions at present, as we move to being [open by default](#) for research outputs. Why is this? The [EPSRC Policy Framework on Research Data](#) has a handy summary. It states that:

"Systematic management and sharing of research data has many benefits for the research community and the public. These include:

- increasing the visibility of research and generating citations, leading to growth of scientific reputation of individual researchers, their research teams and their institutions
- reinforcing open scientific inquiry
- protecting against use of faulty data by allowing published results to be independently verified, refuted or refined, thus improving the overall quality of research and encouraging diversity of analysis and opinion and helping to resolve scientific disputes
- stimulating new approaches to data collection and methods of analysis
- increasing awareness of research in related areas leading to more opportunities for collaboration
- allowing re-use of data for research not foreseen by the initial investigators – this increases the efficiency of use of public funding by avoiding unnecessary duplication of data collection.
- permitting the creation of more highly powered data analysis by combining data from multiple sources
- facilitating education of new researchers and the wider public"

Like many other institutions we are working our way through what these changes mean for us in terms of open access publications, open data, and perhaps more importantly "open working practices" like this blog. With this in mind we thought it would be useful to sound out the wider UK HE community on how they are responding to these new imperatives, as part of our own [Research Data Management](#) project at Loughborough University. Many thanks to the [Digital Curation Centre](#) for helping us to get the message out to people.

We had a good response to the survey, with feedback from the 38 institutions listed at the bottom of this post. We took the decision to publish an anonymized summary of the responses rather than the raw data, to allow people to speak more freely. The authors extend their thanks to everyone who contributed.

2. Executive Summary

The survey results are summarised below, and we have also drawn some personal conclusions:

- Whilst the policy framework has been in place for some time now (e.g. six months for the EPSRC mandate), the survey results show that most institutions are still at the planning stage for the long term archival of their research data.
- Those institutions which had already launched or were preparing a Research Data Service had identified a significant resource requirement both in terms of equipment (typically 0.5PB - 1PB of storage) and staffing (51 fixed and permanent FTE across the 38 respondents) to implement and support the service.
- A small number of institutions had previously benefitted from Jisc funding under the Managing Research Data programme¹ to help develop Research Data policies, advice and even services. It was clear from our engagement with the community over the survey that people were generally aware of and looking to draw upon the outputs of these projects - and also to work with the Jisc funded Digital Curation Centre.

Ongoing funding for equipment and staffing will clearly be crucial to the success of this initiative. We thought it would be instructive to look at a quick summary of the staffing resources (in Full Time Equivalent posts) that the responding institutions have already put into their Research Data Management projects and Research Data services:

<u>Summary</u>	<u>Fixed Term</u>	<u>Permanent</u>
Project Manager	8	2
Data Librarian	4.5	9
IT Support	6.5	4
Research Support	3	2
Other	6.5	5.5
TOTAL	28.5	22.5

If we took a median salary figure of £40K after on-costs, reflecting the specialist skills required to run these services, then this represents a commitment of over £6m per annum from just the 38 institutions listed. It also goes without saying that storage facilities on the Petabyte scale don't come cheap - essentially £500K-£1m in capital costs for the bare minimum viable product, on a four to five year replacement cycle. Perhaps more significantly, this is a whole new area for most people, and there will be a significant learning curve and training requirement - both for support staff and researchers.

We could also observe here that institutions are tending to work on the basis that they need to create their own institutional facility, with all of the consequences for resourcing that this implies. Whilst there is evidence of best practice being shared, institutions could also find ways to collaborate around the actual service provision, thereby reducing their operating costs - and perhaps building upon the regional

¹ <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>

supercomputer centres, such as HPC Midlands². This is a live discussion for the regional clusters of research intensive Universities, but it also has wider applicability to the "long tail" of the sector as a whole.

Once we are all up and running with our Research Data services (whatever form they end up taking) it will be interesting to contrast the setup and running costs with the uptake of the data through downloads, dataset citations, and the evidence of data from UK HE being re-used and repurposed by research and industry. This is an area we recently touched upon in Martin Hamilton's talk on Making e-Infrastructure Accessible to Industry³ for the UK e-Infrastructure Academic User Community Forum.

² <http://hpc-midlands.ac.uk>

³ <http://blog.martinh.net/2013/09/making-e-infrastructure-accessible-to.html>

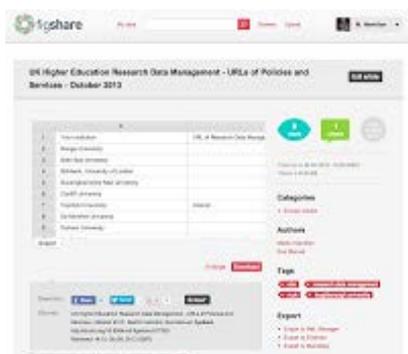
3. Get The Data

The responses which follow below are available to download in raw (redacted) form, hosted by figshare as DOI <http://dx.doi.org/10.6084/m9.figshare.817926>.



[Raw responses \(redacted\)](#)

Many respondents provided links to their institutional Research Data Management policies and (where available) Research Data Services. You can click on the image below to download a spreadsheet of this information, which has DOI <http://dx.doi.org/10.6084/m9.figshare.817925>.



[Links to RDM policies and services](#)

The spreadsheet for our analysis of the data from the survey is also available for download, with DOI <http://dx.doi.org/10.6084/m9.figshare.816938>:

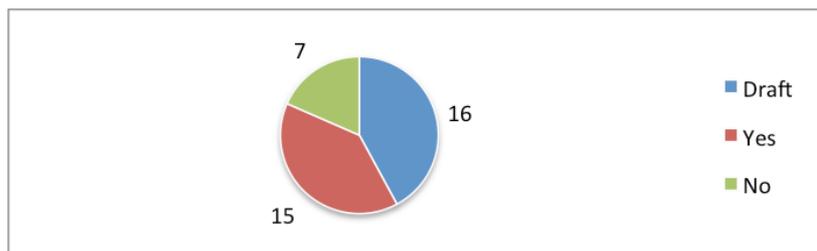


[UK HE RDM Survey – Analysis](#)

4. Survey Responses

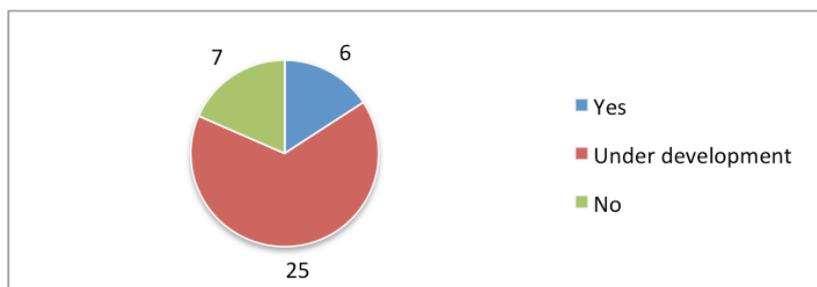
4.1 Does your institution have a Research Data Management Policy?

31 of the respondents (82%) had either a draft RDM policy, or one that had already been approved institutionally.



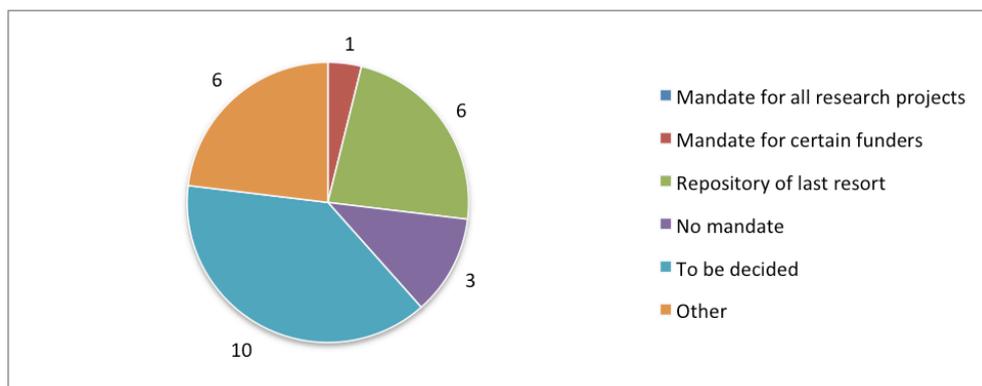
4.2 Does your institution have a Research Data Service?

6 of the responding institutions (16%) had an operational Research Data Service, with a further 25 (66%) under development.



4.3 Scope of Research Data Service

The most common scenario was to use the institutional Research Data Service as a "repository of last resort" in the event that no other domain specific data archives (such as the ESRC funded [UK Data Archive](#)) exist. No respondents had mandated use of their Research Data Service for all research outputs.

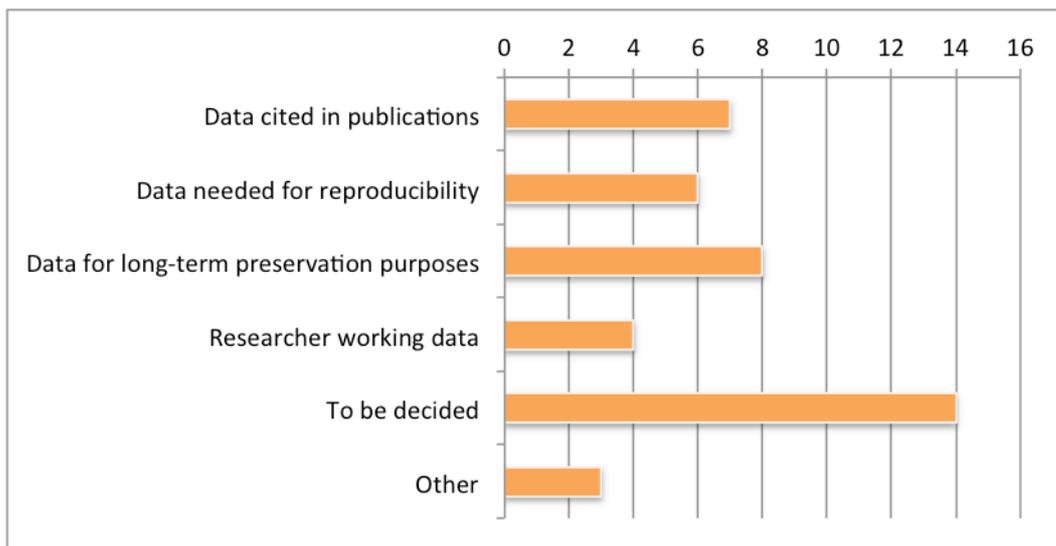


Respondents also stated that:

- We also aim to collect metadata for all datasets
- Faculties will use different elements of the service and this will depend on discipline/funder requirements - the policy gives a framework for this engagement
- We are not providing a service for datasets themselves at the moment, but are gathering metadata
- In theory the repository of last resort but it depends what you mean by 'record all datasets'. We aim in future to record the existence of datasets affiliated to the institution regardless of location (e.g. if they're in UKDA) but don't anticipate housing the data itself in such a case.

4.4 Extent of data to be held in Research Data Service

Respondents could select from multiple categories for this question. The guidance notes with the survey form indicated that "data" as a term should be treated broadly, e.g. any and all working data, data (and potentially software) required to re-run experiments and reproduce results, or just data explicitly cited in publications.

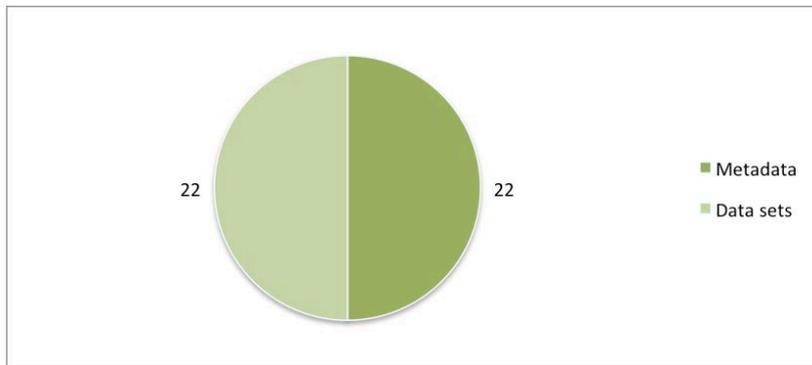


Respondents also stated that:

- We are still discussing researchers' working data and data sharing
- Storage service is the first service, archiving to follow
- Significant data to be deposited or linked to if held elsewhere in line with the policy and disciplinary perspective. We would anticipate that in most circumstances data supporting publications, data that cannot be reproduced will be significant. Etheses formal deposit process just altered to include optional deposit of underpinning data.

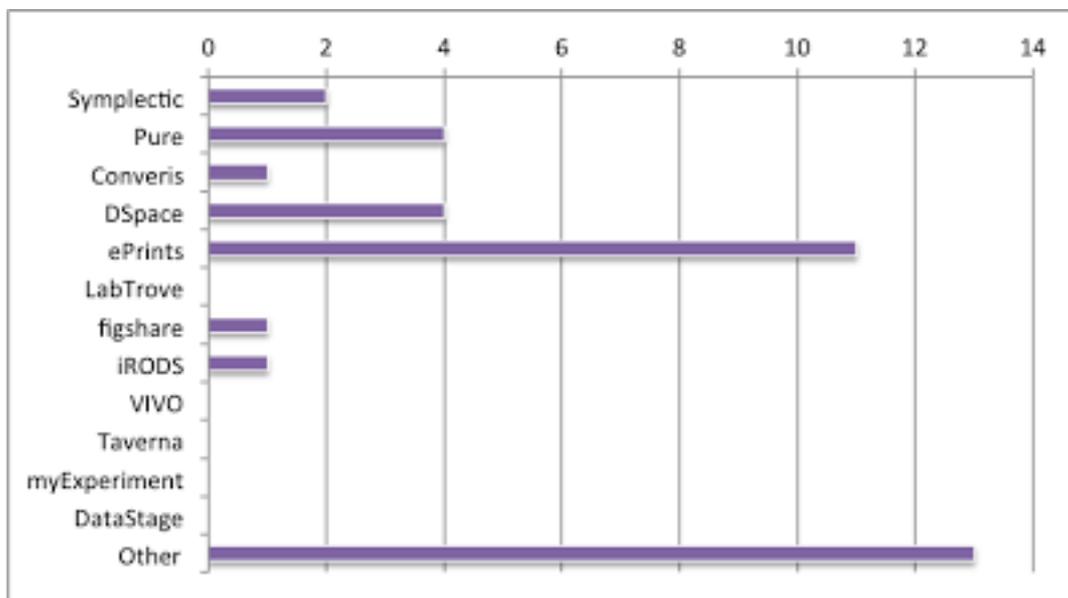
4.5 Categories of information your Research Data Service stores

Most institutions were storing or aiming to provide facilities for both data and metadata (cataloguing information) in their Research Data Service. Two respondents planned to hold only metadata in the central service, and two said that they would hold only data.



4.6 What software does your Research Data Service use?

Respondents could select from multiple categories for this question. A large proportion of respondents intended to use existing Research Information Systems (e.g. Pure, Converis or Symplectic) or institutional repository software (e.g. DSpace, ePrints) for Research Data Management.

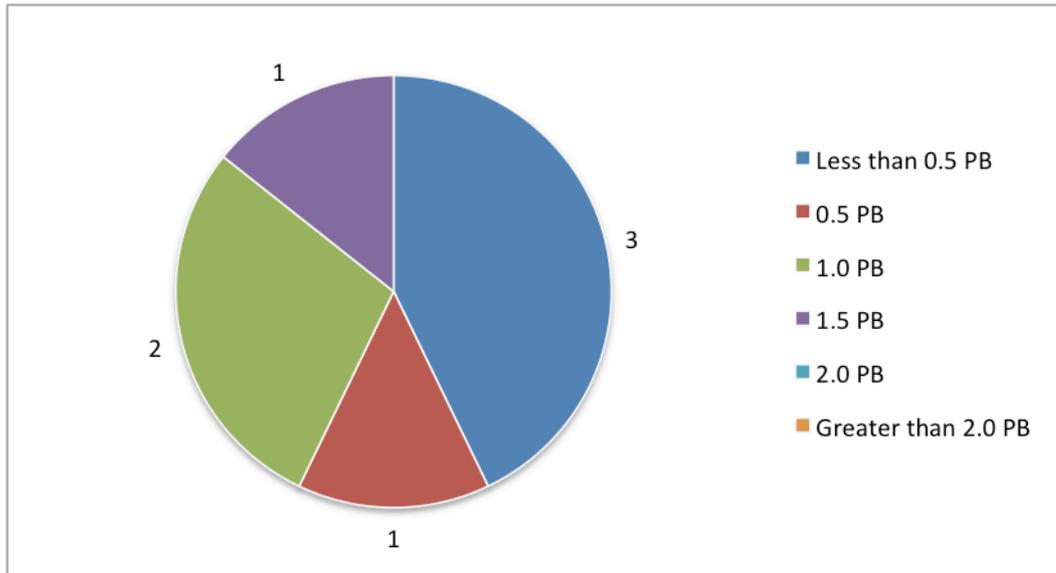


Some further feedback was received on other software products that institutions were using, or planning to use:

- DataBank
- DataFinder
- DMPonline
- Equella
- Fedora/Hydra
- iRODS

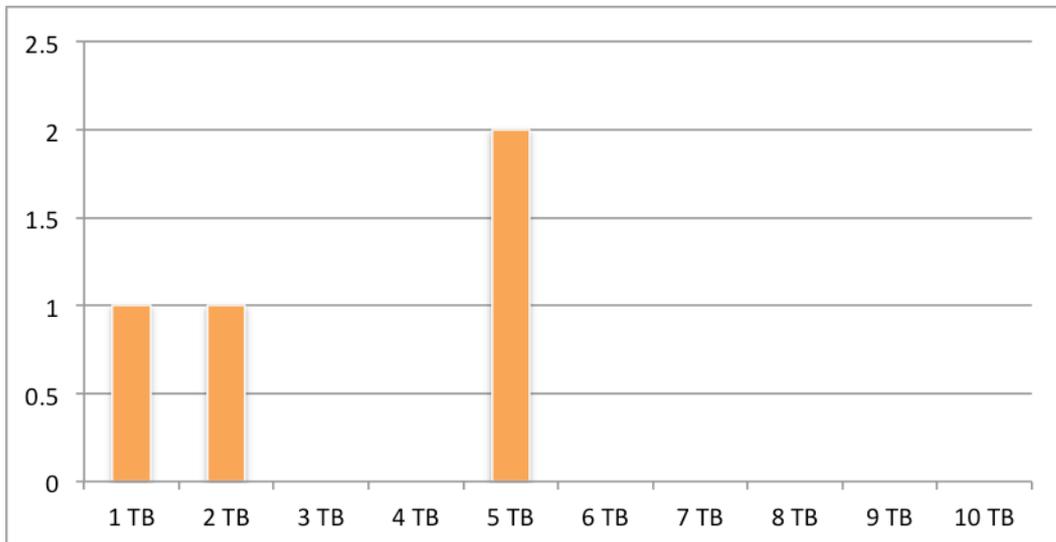
- Open Text Content Server
- Other software and services developed in-house

4.7 How much storage is available for research data sets?



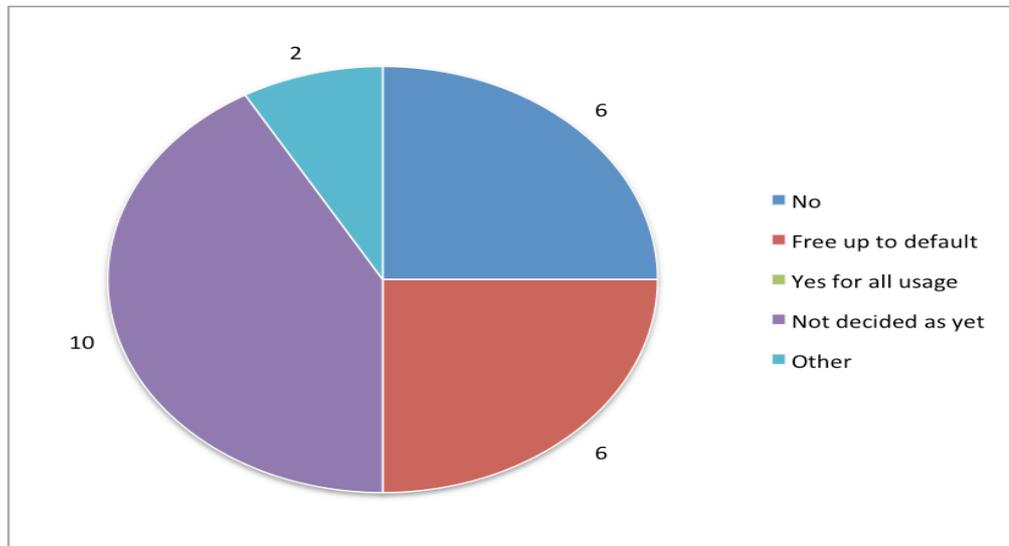
4.8 How much storage is each researcher allocated by default?

Respondents were advised to round up or down to the nearest TB if the scale on the survey form was not an exact match.



4.9 Does your institution charge for the Research Data service?

6 institutions responded that they did not charge (or would not be charging) for their Research Data Service. No institutions said that they would charge for all usage.

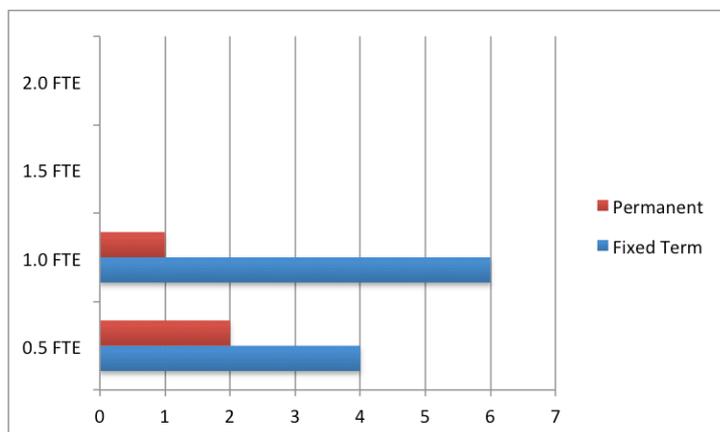


Respondents also stated that:

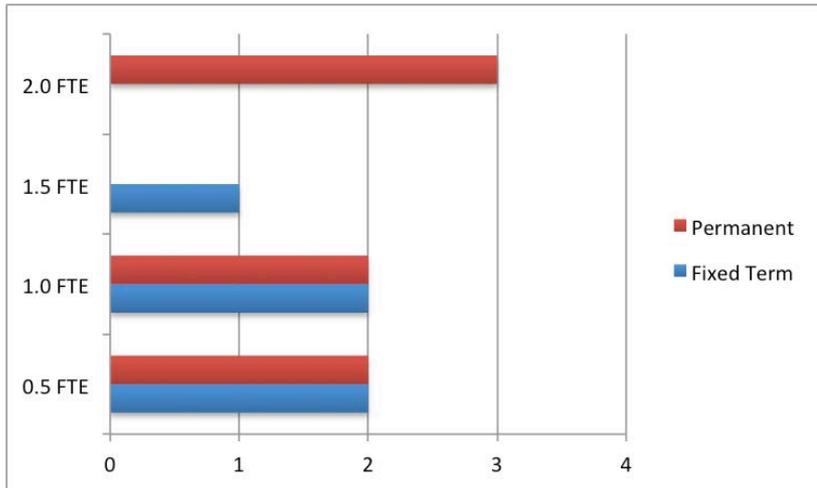
- We don't yet have a cap on total or researcher specific storage capacity
- Faculty allocations and top up charges for working data storage - this is not stored as part of the eprints service which is mainly aimed at completed datasets

4.10 Staffing associated with the Research Data service

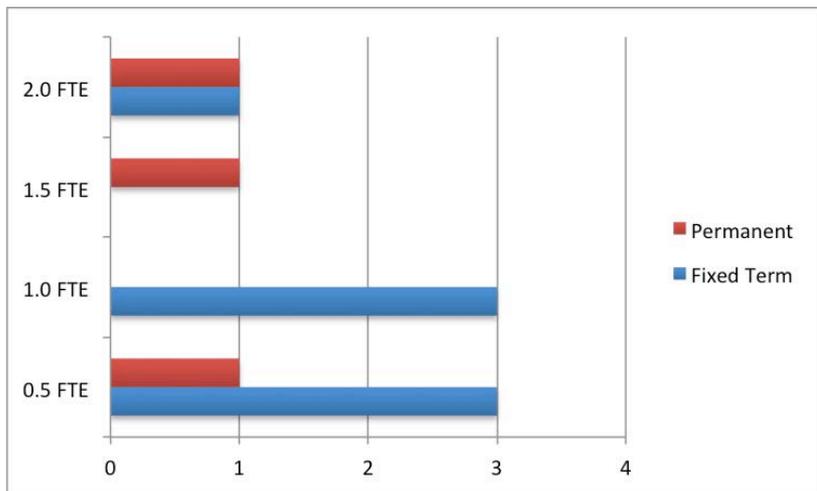
Respondents were asked to quantify the staffing element of their Research Data Management project or service. This was broken down by fixed term contract and permanent staff, and then by role - project manager, data librarian, IT support, research support and "other". The X-axis figures on these charts are the number of institutions which responded with a particular answer, e.g. three institutions had a 0.5FTE fixed term project manager role.



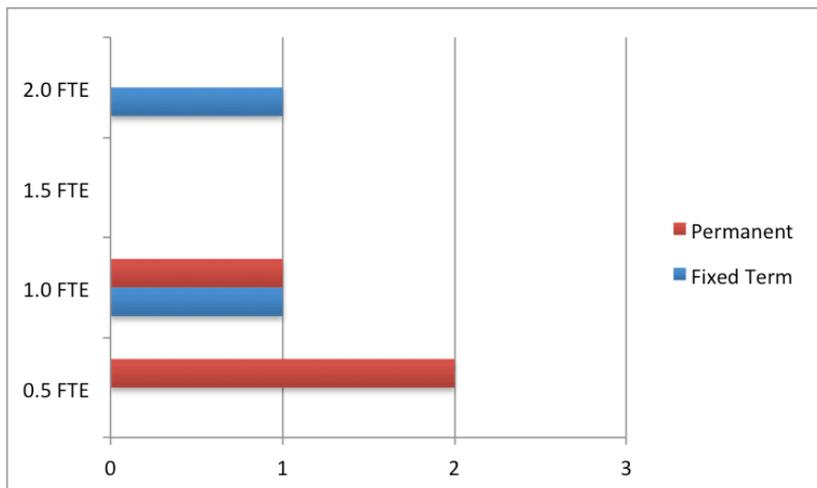
FTE Allocation to Project Manager Role



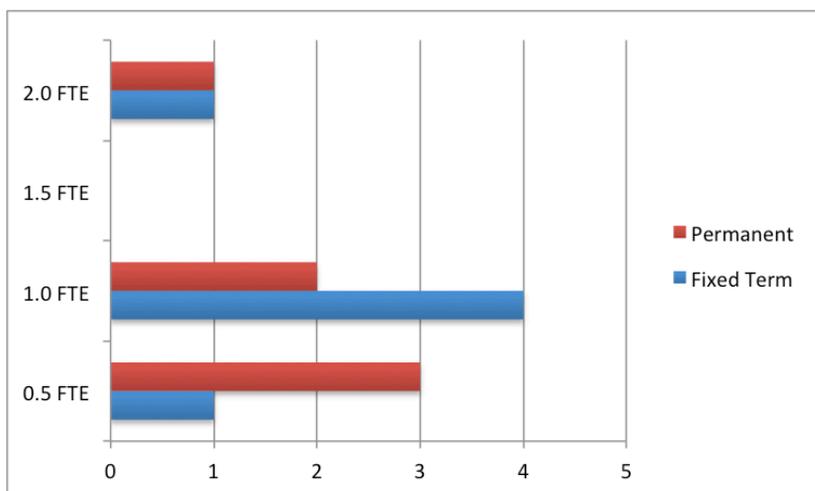
FTE Allocation to Data Librarian Role



FTE Allocation to IT Support role



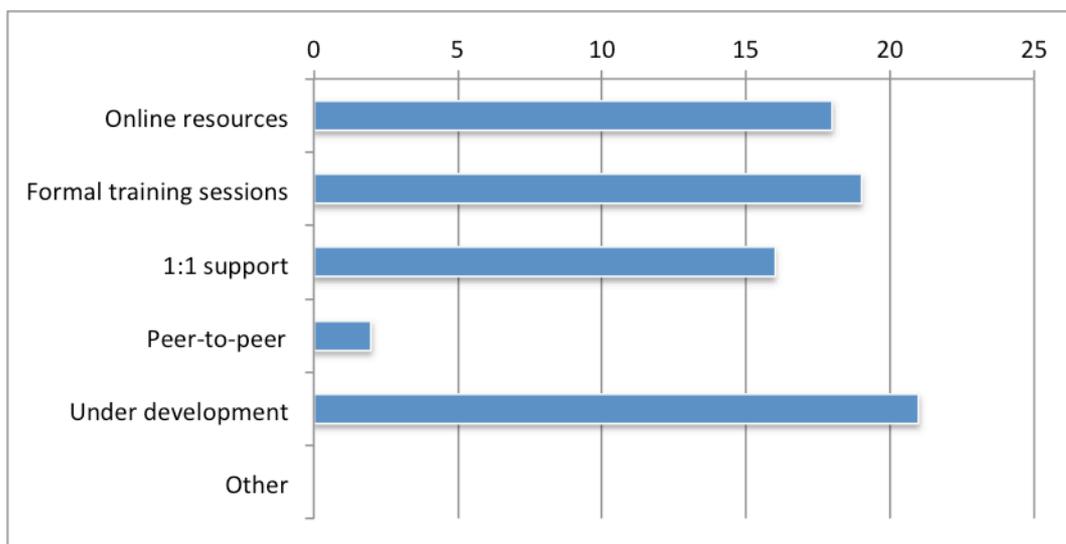
FTE Allocation to Research Support role



FTE Allocation to other roles

4.11 Support available for Research Data Management at your institution

Respondents were asked to quantify the areas where their institution was offering or aiming to offer support for Research Data Management, e.g. online resources, formal training sessions, and so on.



5. Further Information

Respondents were invited to provide additional information about their Research Data Management project or service. 21 out of 38 (55%) institutions contributed to this section.

"The following resource are, or will be provided:

1. One-to-one advice and guidance for projects
2. RDM support site (to be launched in October)
3. RDM training events for staff (recently added to training programme - to be provided in new term)
4. RDM training events for students (currently provided)
5. Web-based RDM training for overseas staff and distance learners (to be developed in 2013/2014)"

"The in-development RDM service at [University name] will be built from a number of tools and services, some of which are in development and some of which already exist (and are not necessarily exclusively concerned with RDM). It's intentionally modular in design with aspects intended to assist researchers meet their data management requirements at various stages of the research life-cycle - it's not simply an institutional repository managed by the libraries. That said, we're still trying to get the resources to actually be able to staff the various parts, so arguably it's not much of an infrastructure at all right now."

"Like others we are in upgrading our offerings and expect we will have more staffing and more storage support within the next year"

"Difficult to answer some of these questions as they assume a certain kind of research data service which includes longer term preservation whereas at [University name] there has been investment in active research data service (although this rapidly filled and is not currently available to new users) but only recently has resource in the Library been funded to pursue longer term preservation other than PhD Theses. Otherwise there have been very significant initiatives in research data management funded through fixed term external resources but this will not be maintained without additional internal investment."

"Please note that 5TB is available for all lead researchers but there is an application process (rather than default allocation)"

"We have split the domain of RDM into three concerns and formed three services around these concerns.

1. Data Storage . Agnostic and commodity storage at a large scale
2. Data Archive. shift in responsibility and concerned with data preservation
3. Search and Access. Enabling the discovery and re-use of data held in the data archive.

These services are interdependent but are being planned and implemented at different stages.

Sitting behind the practical procurement and commissioning is comprehensive effort with data policy, data licencing, and outreach/communication."

"This was quite hard to answer as we see the data service encompassing many elements - data management planning support (now relatively mature), support for the data catalogue and archive which is really only just launched so just getting

started, training sessions and deskside support for PGRs and staff (several sessions run through central graduate centre training, DTC, Faculties) with a model to scale up.

The staffing is all embedded. This is an embedded part of some Research Support Officers in Research and Innovation services, Business Relationship managers in IT and Academic Liaison librarians. In the library we have allocated 3.0 FTE of time to RDM support (Liaison and liaison support) and embedded RDM as a strategic service senior lead as part of my role.

Additional issues currently being investigated include a draft DataCite DOI minting policy and trusted repository benchmarks."

"RDM policies have been written but are not yet public. Details of storage capacity and individual allocation yet to be decided. Details of staff once the repository has been set up yet to be decided"

"We only currently have 0.2 of a project manager."

"[University name] is part way through a procurement exercise to implement a Research Information System which will be used as part of the solution to manage research data and metadata.

A fixed term Research Data Manager and an Information Systems Specialist have been appointed and IT support for RDM is coordinated through the IT Business Partner for Research.

A working group drawn from the Library, Computing and Information Services and the Research Office was formed in October 2012 responsible for the development of policy, support and guidance and reporting to the University Research Committee."

"We are currently developing a training programme for RDM which will be launched in November 2013. Discussions are ongoing with a number of prospective suppliers for data storage and curation software, and other institutions around their approach to RDM."

"Project working group initiated to decide RDM processes and services. Senior representation from Library, IT, researchers and finance. Input from DCC has been initiated."

"Business Case for Research Data and Information Management has been approved by the Steering Group and Business Change Oversight Group and will be presented to the University Executive Board 14/10/13 for investment decision."

"This survey was jointly completed by myself and [name] from the University's Research Office. We have been working on Institutional support for Research Data Management with assistance from DCC. The University has a steering group chaired by [name] - Pro Vice Chancellor Research and Knowledge Exchange. We have conducted a data audit throughout the university - interviewing some 90 researchers of varying roles. We are currently working on providing advice and services to all research active staff. So far [name - University's Research Office] is acting as Research Data Coordinator and we have run a half-day session in the library looking at the possible roles for the library. No resources has specifically been set aside for research data management yet."

"The Research Data Service includes support for research data management (as listed), storage for active research data (currently 1TB per funded project with

charges thereafter), and a research data archive, which is currently under development and will use both EPrints and Pure."

"Scoping Research Data Services this semester"

"The [University name] have identified several key RDM areas that require action. These priority areas were identified by an internal RDM pathfinder project which took place in 2011/2012. The RDM Co-ordinator was appointed in July 2012 in order to take forward these actions and co-ordinate institutional RDM activity amongst key stakeholders and build institutional RDM capacity and capability."

"Research Data Management at [University name] is primarily focusing on the policy, processes, training and metadata capture aspects before tackling the issue of where the datasets themselves should finally rest. We have undertaken quite an extensive gap analysis to look at what needed to be done, and have currently established a framework of guidance (mostly webbased) and training, and have a policy in draft. We are looking into the need for support staff training in research data management at the moment and are working on the requirements for a metadata catalogue. With regards the data itself - at the moment we are mandating that researchers should find a suitable external repository to host their data in the longer term and are not providing space at [University name] apart from as a very very last resort as we don't currently have the capability to curate data over the longer term."

"Some of the questions are difficult to answer definitively because the service is still under development. Although some of the posts are fixed term, they are primarily secondments of permanent staff - so we are building a core of experience and expertise at the institution."

"Aspirational at present. Steering Group in place with draft policy and action plan under way. Identified need for additional resources dedicated to RDM to ensure faster progress. Further ahead with Open Access but recognise the need to link together for greater success."

"We are at the very early stages, but we have a general OA policy for publications agreed and a RDM statement will be following by the end of the year. We will use EPrints for the time being but this has not been assessed as a long term or large scale solution. We are currently managing with existing staff between Information Services and Research Services."

6. Survey Respondents

38 institutions that had responded to the survey by 2nd October 2013. The survey is still open for further responses, and we will update this report should we receive a significant number of additional responses.

Bangor University
Bath Spa University
Birkbeck, University of London
Buckinghamshire New University
Cardiff University
Cranfield University
De Montfort University
Durham University
King's College London
London School of Economics and Political Science
London School of Hygiene and Tropical Medicine
Loughborough University
Newcastle University
Oxford Brookes University
Queen Margaret University
Royal Holloway University of London
SOAS
The Open University
University College London
University of Aberdeen
University of Bath
University of Bristol
University of East London
University of Edinburgh
University of Glasgow
University of Hertfordshire
University of Hull
University of Kent
University of Leeds
University of Leicester
University of Liverpool
University of Northampton
University of Oxford
University of Sheffield
University of Southampton
University of Stirling
University of Strathclyde
University of Surrey