

ESTABLISHING INCENTIVES AND CHANGING CULTURES TO SUPPORT DATA ACCESS

*Report of the Expert Advisory
Group on Data Access*

REPORT ANNEXES

ANNEX A: WEB SURVEY OF RESEARCHERS AND DATA MANAGERS

KEY FINDINGS

1. Survey respondents showed a clear appreciation of the value of data sharing in advancing scientific knowledge. Resistance to data sharing was instead mainly an issue of feasibility and lack of reward.
2. Respondents emphasised that data sharing requires considerable resources, and therefore requires dedicated support from funders. There was also a widespread view that funders weren't effectively tracking the implementation of data sharing plans post-award.
3. Several respondents suggested that institutional cultures generally lack support for and do not expect data sharing. Thus there is little incentive to spend the time required to share data.
4. There was a clear view that there is little reward or recognition for data sharing nor usually any consequences for not sharing data, both in institutions and in the context of grant applications.
5. Many respondents held the view that data managers lack a defined career structure and their role was felt to be undervalued in institutions.
6. Respondents suggested that current infrastructure is inadequate to support the sharing of data. At present most data management and sharing tools are not sufficiently user-friendly, and researchers are not well trained to use them.
7. There was a lack of awareness and use of data citation metrics. Therefore there is little reason for citation of datasets, despite the potential recognition which could be attained.
8. Respondents made a range of suggestions for how incentives for data sharing could be enhanced, focusing in particular on three key themes:
 - **Funding:** Funders should make increased dedicated funding available for data sharing, and make it an explicit requirement of funding.
 - **Culture & recognition:** Recognition of data sharing should be increased, including through the Research Excellence Framework (REF), and through increasing the status of data managers and making their career structure more defined.
 - **Infrastructure:** Infrastructure for data sharing should be improved, enhancing accessibility and ease of use. Technical support in data sharing should be provided for researchers.

AIMS AND BACKGROUND

The survey was part of an Expert Advisory Group on Data Access (EAGDA) project to examine the issues surrounding research data access and sharing, with the aim of identifying potential ways to incentivise scientists to share data. As part of this project, we created a web-based survey to gauge the views of both researchers and data managers at different stages in their careers on the current barriers to data sharing, and potential ways in which incentives could be enhanced.

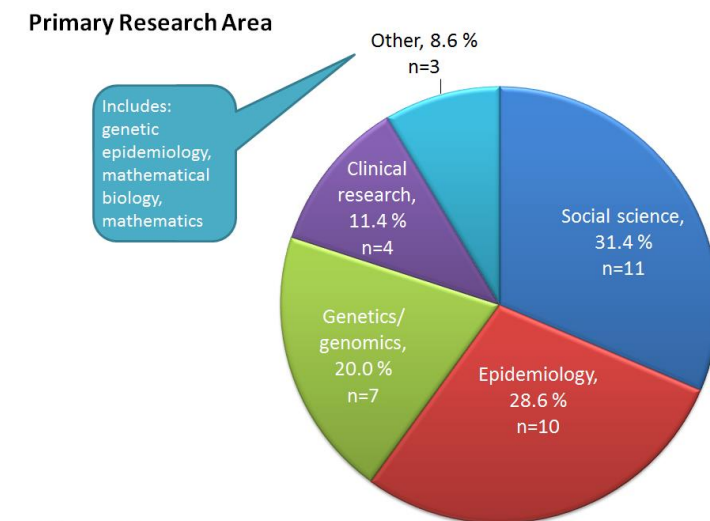
METHODOLOGY

The web survey (see **Appendix**) was sent to 110 nominated individuals across four funders: the Trust, MRC, ESRC and Cancer Research UK. Each funder was invited to suggest up to 30 individuals from a variety of data-intensive fields and career stages. There was difficulty in identifying post-doctoral students and data managers from funder information, so those in receipt of the survey were encouraged to forward it on to members of their research teams and other colleagues.

The survey was in the field from 27 June to 19 July 2013, during which time 35 responses were received. Respondents included researchers across the fields of genetics, epidemiology and social science, at a mix of research positions (see **Figures 1 and 2**).

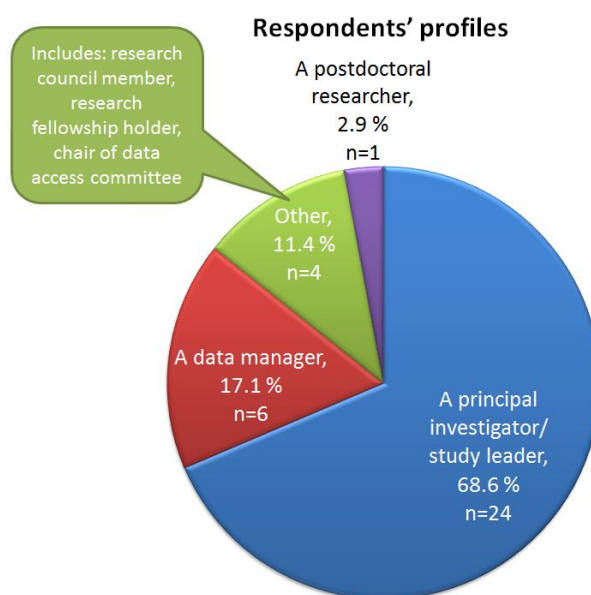
Figure 1

The respondents held varied positions within scientific research. The majority were PIs (n=24), and some were data managers (n=6) (see **Figure 2**). Unfortunately, only one response from a post-doctoral researcher was obtained.



Base: All respondents n = 35
Q: Which of the following best describes your primary research area?

Figure 2



Base: All respondents n = 35
Q: Which of the following best describes you?

The survey included questions on the following topics:

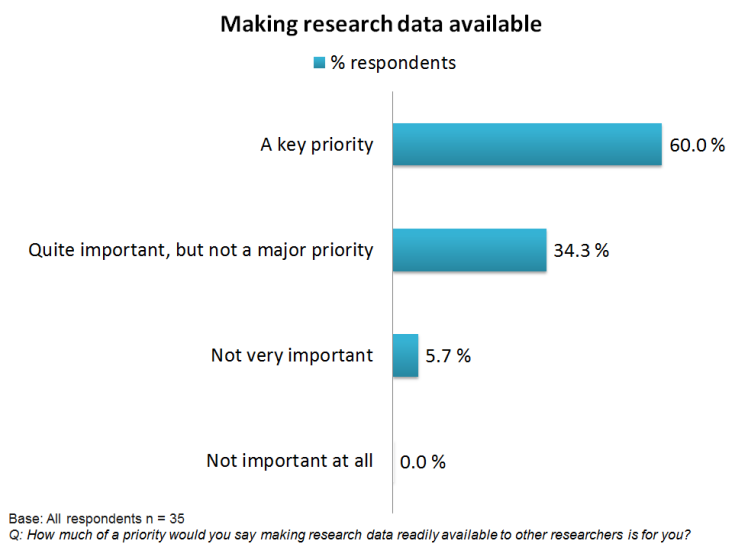
1. The value of data sharing
2. Barriers to data sharing
3. Funding agencies
4. Research institutions
5. Career development
6. Multi-author publications
7. Publication moratoria
8. Dataset metrics
9. Recognition
10. Incentives suggestions

RESULTS

1. The value of data sharing

The majority of respondents showed a clear appreciation of the importance of data sharing, with 60% saying that sharing their data is a key priority (see **Figure 3** and **Box 1**).

Figure 3



Box 1: Value of data sharing – examples of comments

“Personally, it's very important.”

“We consider sharing our data to the wider scientific community is vital to enable collaborative research and advance scientific knowledge and dissemination.”

“I think it's important for science to share as much as we possibly can. Many researchers forget that they should be in it to further knowledge, not simply to advance their careers.”

“We are keen for as many researchers as possible to share the valuable data that we collect.”

“It is essential that the detailed and rich data that we collect on longitudinal studies is made available to the wider research community. There are such a wide range of research questions that can be addressed by the data that we collect and manage that it is imperative to share it as widely as possible to ensure the maximum scientific gain.”

“We use public funds to amass large data sets. Their value lies in their exploitation.”

“For science to advance, for research questions to be answered, data must be made available to legitimate researchers.”

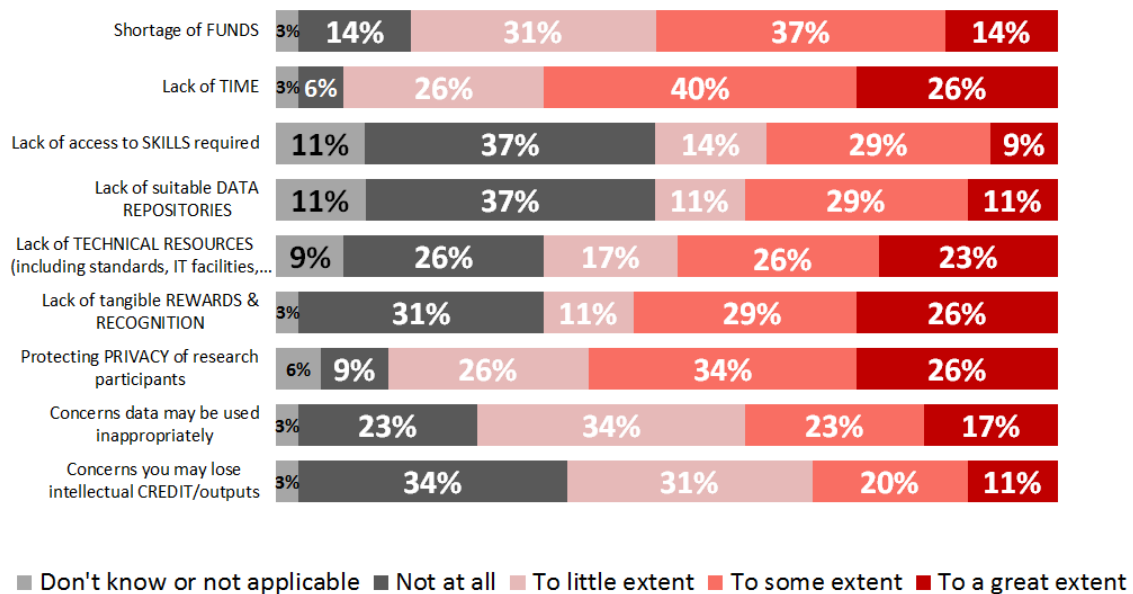
2. Barriers to data access and sharing

It was noted that collaboration involving data exchange between groups is becoming increasingly common. However, respondents highlighted a considerable number of issues associated with data sharing.

The extent to which the different factors were barriers to data sharing for each researcher was highly varied. Several respondents found the following factors to be particular obstacles: lack of time (where 66% indicated it constrained them to a great extent or some extent), funds (51%), recognition (55%) and technical resources (49%), as well as the need to protect the privacy of research participants (60%) (see **Figure 4**).

Figure 4

Barriers to data sharing



Base: All respondents n = 35

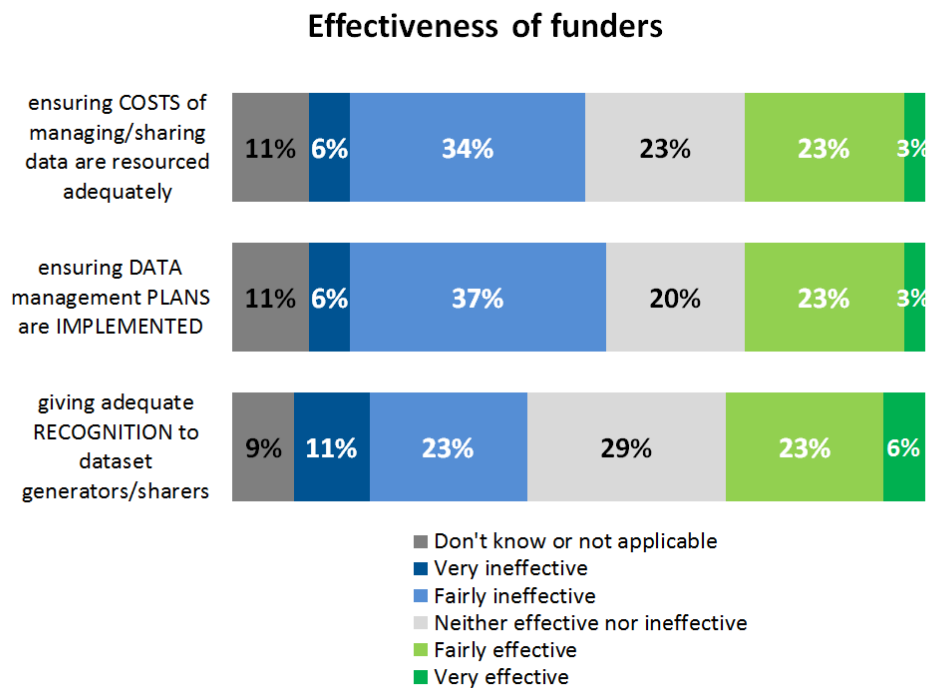
Q: Thinking of the research team/environment in which you work, to what extent do the following factors act to constrain you/your team in making datasets more widely available?

3. Impact of funding agencies

Many respondents felt that data sharing is not effectively integrated into the funding process. Although grant applications include a section on data management, it was perceived that this is not given adequate weight by some applicants and reviewers. It was noted that funders are increasingly asking for more detailed data management and sharing plans, which may help to tackle this. However, it was highlighted that no processes exist for supporting the tracking and further development of data management plans post-award.

Several respondents viewed funders as not being effective at financing data sharing costs, ensuring implementation of data sharing plans or in assigning due recognition (see Figure 5).

Figure 5



Base: All respondents n = 35

Q: Overall, in your experience, how effective, or not, are funding agencies in terms of supporting the following regarding their policies on data access and sharing?

Box 2 – Support from funding agencies for data management

“In my experience most grant applications contain perfunctory statements about data release [that] say that policies will be followed. Details beyond such a statement are almost never provided. To my knowledge no researcher has ever suffered any detrimental impact from failing to share data either as a reduction to a current grant or any limitation to a future grant.”

“Although the funders that I know best now expect a statement about data sharing, this is usually regarded by investigators as an irritating extra box on application forms rather than a major part of the work they propose, and hence I suspect is often not taken seriously.”

“Getting funding for data management is difficult. Funding agencies may cite it as a priority, but funding committees seem to think differently.”

“Making data accessible and usable takes time and scientific expertise as well as technical skills. Research grants don't usually resource this adequately despite there being an expectation by funders to make data accessible.”

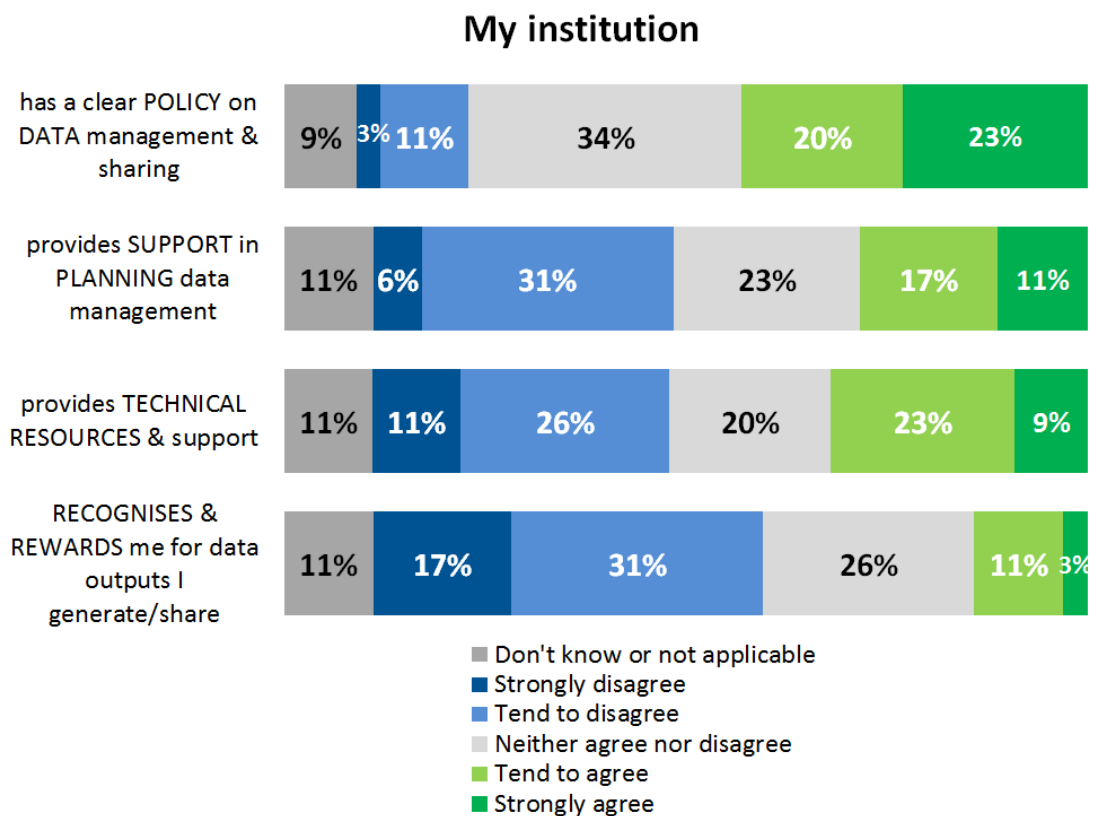
“None of the funders have as far as I am aware a process for supporting the development of and explicit review of data management plans at any stage in the research data lifecycle. Some research contracts specify data sharing and data deposition plans but there is no follow through in terms of the real support that is required.”

4. Culture of institutions on data sharing

Though some institutions were described as being encouraging of data dissemination, many respondents viewed the culture of their institution as neither expectant nor supportive of data sharing. Only 43% of respondents agreed that their institution has a clear data management policy, and many did not appear to be sure (see **Figure 6**). This limited awareness may indicate that certain institutions have policies in place, but these may not be sufficiently well publicised.

The minority of respondents agreed that their institutions provided support in creating data management plans (28%) or technical resources (32%). Likewise, only 14% agreed that they are given clear recognition for their data outputs (see **Figure 6**).

Figure 6



Base: All respondents n = 35

Q: To what extent do you agree or disagree with the following statements: My institution...

Box 3 – Support from research institutions for data management

“My institution is supportive of the principle, but I think has been relatively neutral on implementation and mechanisms for most disciplines.”

“Research interest and supportive culture but difficult to obtain any additional technical support for specific projects”

“There is a policy in place on data sharing however I have never been made aware of the principles of this policy.”

“The UK university rewards structure does not treat data collection and dissemination kindly in terms of kudos and incentives. There are also questions, which we are grappling with, about the ubiquity of research skills (specifically quantitative and methods skills)”

“There is a lack of data management skills in the wider research community and many people find it difficult to work with complex longitudinal files. With additional resources we could provide further training in data management and more 'user-friendly' datasets.”

“Being able to have data that stands alone without long term support is a concern as we want to ensure that researchers understand the data collection, its context and how we may have transformed it to make it more usable”

5. Career development

Respondents were asked whether issues associated with data sharing may deter researchers and data managers, particularly at an early career stage, from pursuing data-intensive fields.

When asked to identify the major issues that constrain career development, respondents commonly cited funding and a lack of job security for both early-stage postdoctoral researchers and data managers. Lack of academic recognition was also a common factor, particularly for contributions to multi-author publications (see section 6).

The work of data managers was also described as being poorly acknowledged (see **Box 4**). Additionally, the career path of data managers was frequently reported to lack a defined structure. The role was reported to lack prestige, reflecting the lack of focus on data sharing as a priority.

Only half (n=3) of the data managers questioned tended to agree that they intended to remain in their current research area, with two respondents tending to disagree, and one abstaining. By contrast, 92% of principal investigators (n=22 of 24) tended to agree or strongly agreed that they planned to stay in their current field.

Box 4 – Career prospects for data managers

“There is no career progression for data managers.”

“No defined career structure at all.”

“Lack of recognition of their skills & lack of a clear career path - can feel exploited by researchers.”

“Although I have held my current position for the past 7 years, I have had no salary increases or promotion for the past 5 years, and been told that there is no scope for progression because I am not considered “academic” staff. My position as a data manager is considered menial, and expendable.”

“Our role appears to be considered menial, as though we were data entry clerks, despite it requiring (among other things) high levels of programming skills and the responsibility of ensuring that the data is anonymised before issue.”

“Data management is not perceived as a professional skill.”

“Data managers do not receive adequate recognition for the work that they do. Researchers assume that data management comprises of editing and inputting data.”

6. Multi-author publications

Data-intensive research often involves multiple contributors, leading to multi-author publications. This minimises the proportion of scientists who can be listed as a main author. Consequently, 74% of respondents indicated that in their field there is difficulty in assessing contributions of individuals to multi-author papers, and 65% said this formed (to some or to a great extent) a barrier to career advancement. Data managers in particular were reported to receive little or no recognition on publications.

This was highlighted as an issue for assessing achievement in the REF, and can have significant implications for career progression (see **Box 5**).

Box 5 – Effect of multi-author publications on career progression

“My multi-author publications in Journals such as Nature Genetics have been considered inadmissible for the forthcoming REF return exercise even though I made a significant contribution. This has significant implications for my career progression.

I believe my international work and collaboration is undervalued by my organisation. This is disappointing as I think consortia are key to making true scientific progress especially in rarer diseases. Unfortunately my current experience has strongly dissuaded me from such a level of involvement in the future.”

Suggestions of how these issues could be overcome included the compulsory specification of each author’s contribution (n=4), or displaying authors’ names within a classification system, such as a hierarchy of significance (n=2), rather than a simple list.

7. Publication moratoria

Publication moratoria were viewed favourably by most respondents. Of those with experience of the use of publication moratoria for datasets (n=15), 80% regarded them to be fairly or very effective in enabling timely access to data whilst preserving data generators' rights and ensuring they receive due credit (see **Figure 8** and **Box 6**). Of the 13% (n=2) who considered moratoria to be very ineffective, one respondent suggested that moratoria could be used to pressurise smaller research groups to share their work, often prematurely.

Of those without experience of their use, 55% tended to agree/strongly agreed that publication moratoria would have value in their field. Eleven per cent (n=4) strongly disagreed, all questioning the need for any data dissemination delays at all.

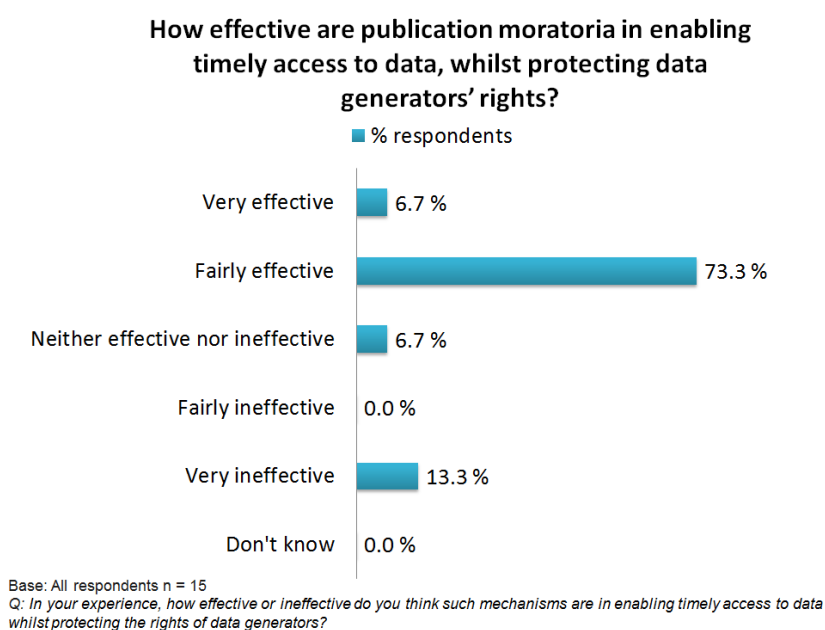


Figure 8

Box 6 – Publication moratoria and preserving data generators' rights

“Losing credit for data is not of consequence if original collection of data is available widely after a suitable period of embargo to allow researchers to maximise their use from the sponsored collection.”

8. Use of metrics

Eighty-three per cent of respondents already tracked the use of their datasets, primarily through details of publications generated using the data (71%), the number of researchers accessing the data (51%) and the number of data downloads (43%).

However, awareness, let alone use, of research metrics relevant to datasets was limited. For each of the four types of metrics considered, 40% or more of respondents were unaware of them, and less than one in five had actually used them (see **Figure 9**).

Figure 9

Metric	Unaware (%)	Use (%)
Permanent researcher identifiers e.g. ORCID	43	17
Dataset citable identifiers e.g. DataCite Dols	54	11
General data resources e.g. Dryad and FigShare	77	6
Publications that describe datasets e.g. Nature Scientific Data	40	11

9. Recognition of data sharing

Only a minority of respondents agreed that data sharing is recognised substantially relative to publications. The vast majority felt that data outputs should be recognised to a greater extent in relative terms (see **Figure 7**).

Figure 7

Context	Current recognition (% respondents who viewed data is recognised to some extent or to a great extent , relative to publications)	Due recognition (% respondents who viewed data should be recognised to some extent or to a great extent , relative to publications)
Grant applications	26	91
REF	11	86
Career advancement	11	86

10. Incentives for data sharing: suggestions

There was a high degree of commonality in the suggestions offered by different respondents to how incentives could be enhanced. The main recommendations fell under three main areas: funding, culture and recognition, and infrastructure.

Funding

A common suggestion was to increase funding for data sharing (n=5). Equally frequent was the suggestion that data sharing should be made a requirement of funding (n=5), and one respondent explicitly advocated sanctions for non-compliant individuals, such as withholding final grant payments or future grants. Two respondents also suggested the inclusion of a more explicit section on data sharing in grant applications.

Box 7 – Funding suggestions

“Clearer commitment to data sharing principles as a requirement of funding”

“Funding research in concert with data collection”

“Check data accessibility according to previous grant commitments for published data as a condition of final grant payments/new grant awards”

“Institutional data storage, with expectation of use, and earmarked funding within grant applications”

Culture & Recognition

Increasing the recognition of data leadership was a frequent suggestion (n=7), with suggestions that data sharing should be viewed as more important relative to publications (n=2), and the REF should recognise data sharing more (n=3). Other respondents promoted the creation of more defined career paths for data managers, and to give them higher status (n=3).

Box 8 – Suggestions for increasing recognition

“Recognise data generation as an output”

“We need a recognition scheme that privileges publication less and, instead, recognises peer reviewed achievement in the construction and dissemination of data.”

“The Research Councils should press to recognition of ‘data leadership’ and seek rewards within the HE system for those who manage large data enterprises.”

“Recognition on a par with publication in peer-reviewed journals”

“Give data managers more status within research institutes”

“Support data manager posts and career progression mechanisms”

Infrastructure

The suggestion most frequently made was to improve the infrastructure for data sharing, enhancing accessibility and ease of use (n=8). One respondent advocated the direct funding of data repositories by funders. There were also further suggestions along the same lines. It was proposed that technical support or training for data sharing should be provided (n=5) and that datasets should be properly cited (n=2), with encouragement of institutions to increase citation (n=1). Improvement of metrics (n=1) and associated training on use of metrics for tracking data outputs (n=1) were suggested.

Box 9 – Improving data sharing infrastructure

“Make data deposition and access easier”

“Support for better tools to generate and access documentation in a structured format, to allow research groups to deposit accessible high quality data”

“Funders to provide central data repository resources”

“More direct technical support to provide highest quality solutions”

“Promote good practice in data sharing through training and education”

“Archives should have a more pro-active role in data management capacity building, to improve the quality of material that is deposited, and also raise the level of what is expected of depositors.”

“Greater institutional engagement in increasing use of citation mechanisms such as data Dols”

“Work together to develop and promote metrics for data sharing”

DISCUSSION

There was a clear appreciation of the value of data sharing amongst survey respondents. In practice, however, data sharing requires time and effort, and therefore financial resources, a shortage of which was commonly listed by respondents as a key barrier. There was a view that funders needed to give greater weighting to data management plans proposed in grant applications and ensure the associated costs were met. If nothing else, these steps would reinforce the message that funding agencies have serious intentions regarding data management.

However, this alone is insufficient to ensure implementation of data management plans. There is great potential to improve the current data sharing infrastructure, both in terms of user-friendliness and accessibility, which are lacking according to many respondents. Funding agencies could directly fund data repositories to ensure this. Furthermore researchers would benefit from training in how to utilise such tools. Institutions have a key part to play in this, since they provide a base for scientists to learn. Data management training could be included as part of the research training provided to scientists at an institution, at minimum in data-intensive fields.

In order to fully integrate this approach into institutions, they could be encouraged to develop a clear policy on data sharing, and to provide greater technical support for data management. Perhaps data managers could be given some responsibility in training researchers, a role which could aid the

establishment of their career structure, which is currently undefined. Increasing researchers' knowledge of data dissemination in this way would reduce its associated burden of time and effort, and could help to tackle the low status of data managers in institutions. The culture around data sharing cannot be altered rapidly or simply, but by providing the necessary facilitative environment, a supportive culture is more likely to develop.

Nevertheless, time-pressured researchers will still require effective incentives to make the effort to share their data. As suggested by a number of respondents, sanctions for those who do not fulfil data sharing requirements could include the withholding of final grant payments or follow-on grants. It will be very important, however to strike the correct balance between rewards and sanctions, as there is a risk that excessive sanctions could create extreme resistance to the data sharing measures, at the expense of efforts to create a supportive culture. Currently there is already some resentment towards funders regarding their insistence of data sharing, so exacerbating this is a real possibility. It may be preferable to improve the infrastructure for data management and increase rewards first, and only subsequently introduce sanctions, by which point they will likely provoke less hostility.

Data sharing can be most effectively rewarded by increasing recognition of data sharers, which respondents have highlighted as a key obstacle. With respect to the issues associated with multi-author publications, the issuing of a contribution statement by each author was promoted in survey responses and is already in place in certain journals. Making this system more widespread may help to address the problem. A hierarchy of contribution significance may not be adequate, as significance is a highly subjective measure. Publication moratoria were also for the most part viewed as being effective at allowing timely data dissemination without compromising the credit assigned to data generators. Thus moratoria could be trialled in disciplines yet to introduce them, with the length of time of a moratorium dependent on the nature of the research.

There is considerable scope for increasing recognition through the development and use of dataset citation, and corresponding metrics. If a researcher makes their dataset available, and it is subsequently used in significant research by other scientists, due recognition can only be assigned if the data is cited. Therefore improvement of metrics used to track datasets and efforts to raise awareness of metrics amongst researchers (which the survey indicated was currently very limited) could be effective at creating an incentive to label one's data for tracking. In order to drive this development, dataset citation could be viewed on a level closer to publication citation, particularly in the context of the REF, which is powerful at promoting whatever factors it takes into account.

The scientific community recognises the need to maximise the benefits of research outputs, including data, and sharing data is a valuable means to strive to achieve this. There is further work to be done to successfully facilitate and encourage data sharing, particularly in the context of funders and the research community at institutions. However, as indicated by the survey responses, there are a number of routes which the scientific community can take in order to progress towards a culture supportive of data sharing, where data produced as a result of the hard work of researchers can be utilised to its full potential.

APPENDIX– Data Sharing Incentives survey

Incentives and culture change for data access and sharing: seeking the views of our communities

As a part of the workplan of the Expert Advisory Group on Data Access (EAGDA), the MRC, ESRC, Wellcome Trust and Cancer Research UK are seeking the views of our communities on ways in which we help researchers address the costs and barriers associated with making research data more widely accessible, and to help foster a culture that supports and incentivises the generation and timely sharing of high quality data resources.

We would be enormously grateful if you could provide your views on these issues, by completing this short survey. The survey should take around 15 minutes your time to complete.

If you have any queries about the survey or the project, please contact David Carr at the Wellcome Trust (d.carr@wellcome.ac.uk). If you experience any technical difficulties with the survey, please contact surveys@wellcome.ac.uk.

Please be assured that all of your answers are strictly confidential and responses will be reported in the aggregate only.

Primary research area

Which of the following best describes your primary research area?

Please select one option

- Genetics/genomics
- Epidemiology
- Social science
- Clinical research
- Other (please specify) _____

Your role

Which of the following best describes you?

Please select one option

- A principal investigator/study leader
- A postdoctoral researcher
- A research student
- A data manager (providing specialist support in data management, curation or analysis)
- Other (please specify) _____

Your research group

Which of the following best describes your research group's approach to data sharing?

Please select one option

- We make most of our research data available – either openly or to other researchers on request
- We share data in the context of research collaborations, but not with other external groups
- We do not generally share research data outside of our research group
- Other (please specify) _____

Making research data available

How much of a priority would you say making research data readily available to other researchers is for you?

It is:

Please select one option

- A key priority
- Quite important, but not a major priority
- Not very important
- Not important at all

Please briefly explain your answer.

Please enter your comments in the box below

Key barriers to data access and sharing

Thinking of the research team/environment in which you work, to what extent do the following factors act to constrain you/your team in making datasets more widely available?

Please select one option per row

	To a great extent	To some extent	To little extent	Not at all	Don't know or not applicable
Shortage of funds	?	?	?	?	?
Lack of time	?	?	?	?	?
Lack of access to skills required	?	?	?	?	?
Lack of suitable data repositories	?	?	?	?	?
Lack of technical resources (including standards, IT facilities, etc)	?	?	?	?	?
Lack of tangible rewards and recognition	?	?	?	?	?
Protecting privacy of research participants	?	?	?	?	?
Concerns that the data may be used in an inappropriate manner	?	?	?	?	?
Concerns that you may lose intellectual credit and outputs	?	?	?	?	?

Please briefly explain the ratings you have given above.

Please enter your comments in the box below

Any other barriers

Please describe any other barriers that exist which constrain you/your team in making datasets more widely available.

Please enter your comments in the box below

Funder support for data access and sharing

Overall, in your experience, how effective, or not, are funding agencies in terms of supporting the following regarding their policies on data access and sharing?

Please select one option per row

	Very effective	Fairly effective	Neither effective nor ineffective	Fairly ineffective	Very ineffective	Don't know or not applicable
Ensuring investigators plan their approach for managing and sharing data as an integral part of funding applications	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ensuring that the costs of managing and sharing data are resourced adequately	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Helping to establish clear expectations for data sharing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Helping to establish clear timescales for data sharing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ensuring and enforcing good practice among data users	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ensuring data management plans are implemented in practice	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Giving adequate recognition to those who generate and share high quality datasets	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please briefly explain your ratings, specifically focusing on any significant differences between funding agencies.

Please enter your comments in the box below

Institutional support for data access and sharing

To what extent do you agree or disagree with the following statements:

My institution:

Please select one option per row

	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree	Don't know or not applicable
...has a clear policy on data management and sharing	?	?	?	?	?	?
...actively encourages me to make research data available	?	?	?	?	?	?
...provides support in planning my approach to data management	?	?	?	?	?	?
...provides the technical resources and support I require	?	?	?	?	?	?
...recognises and rewards me appropriately for the data outputs I have generated and shared	?	?	?	?	?	?

Please briefly explain your ratings, focusing on any issues you see as particular barriers.

Please enter your comments in the box below

Publication moratoria

In some research fields, protected periods exist (sometimes termed publication moratoria) where access to datasets is provided on the condition that the data generator's right to first publication is respected.

Do you have experience of such systems in the context of your own work?

Please select one option

- Yes, I have experience of sharing or accessing data under such terms
- No, I do not have experience of sharing or accessing data under such terms

Publication moratoria

In your experience, how effective or ineffective do you think such mechanisms are in enabling timely access to data whilst protecting the rights of data generators?

Please select one option

- Very effective
- Fairly effective
- Neither effective nor ineffective
- Fairly ineffective
- Very ineffective
- Don't know

Please briefly explain your answer.

Please enter your comments in the box below

Publication moratoria

To what extent do you agree or disagree that such mechanisms have/would have value in your field?

Please select one option

- Strongly agree
- Tend to agree
- Neither agree nor disagree
- Tend to disagree
- Strongly disagree
- Don't know

Please briefly explain your answer.

Please enter your comments in the box below

Training and career development for data management and sharing

Thinking about your own career, to what extent do you agree or disagree with the following statements?

Please select one option per row

	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree	Don't know or not applicable
Adequate training opportunities are available to me to develop the skills I need to manage and share data	?	?	?	?	?	?
I intend to remain in the research area I work in currently	?	?	?	?	?	?
My contribution and achievements are recognised	?	?	?	?	?	?
Looking ahead, I can see a clear career path for myself in the academic research sector	?	?	?	?	?	?

Please briefly explain your ratings.

Please enter your comments in the box below

Issues that constrain career development

In general terms, and very briefly, what would you say were the major issues (if any) for your research field that constrain the career development of:

Please enter your answers in the boxes below

(i) Early-stage postdoctoral researchers _____

(ii) Data managers _____

Multi-Author publications and attribution

One particular issue that has been flagged in data-intensive fields is that research increasingly involves large consortia, and resulting papers can have large numbers of authors – meaning that the contribution of individuals is hard to determine.

Does the above issue emerge in the context of your own research?

Please select one option

- Yes
- No

Career advancement in your own field

In your opinion, to what extent does this constitute a barrier to career advancement in your own field i.e. in terms of recognition of a researcher's contribution to a paper?

Please select one option

- To a great extent
- To some extent
- To little extent
- Not at all
- Don't know

Please briefly explain your answer.

Please enter your comments in the box below

Measuring the value of research data outputs

Which of the following types of information do you/your team collect to track use of the datasets you generate and share?

Please select all that apply

- Number of data downloads
- Number of researchers accessing the data
- Details of publications generated using the data
- Case studies of impacts resulting from secondary use of the data
- Other (please specify) _____
- None of the above

Emerging mechanisms and metrics

Do you use any of the following emerging mechanisms and metrics?

Please select one option per row

	I am aware of, and use this metric	I am aware of, but do not use this metric	I am not aware of this metric	Not applicable
Permanent researcher identifiers (e.g. ORCID)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Attribution of permanent and citable identifiers for datasets (e.g. DataCite DoIs)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
General data resources to enable sharing of data underlying published outputs (e.g. Dryad and FigShare)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Publications that describe datasets (e.g. Nature Scientific Data)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please list any other emerging tools that you are aware of that enable researchers to track (or potentially track) the secondary use of data.

Please enter your comments in the box below

Value of emerging mechanisms and metrics

Overall, to what extent do you believe these tools, as a whole, will have value in helping researchers measure and demonstrate the use of their data?

Please select one option

- To a great extent
- To some extent
- To little extent
- Not at all
- Don't know

Please briefly explain your answer.

Please enter your comments in the box below

Recognising the value of data outputs

Relative to publications, to what extent do you think generation and sharing of high quality datasets are currently recognised in assessing a researcher's outputs and performance?

Please select one option per row

	To a great extent	To some extent	To little extent	Not at all	Don't know
Assessing a researcher's track record in the context of grant applications	?	?	?	?	?
Making decisions in career advancement at institutional level	?	?	?	?	?
Assessing researcher contributions in the context of the REF	?	?	?	?	?

Recognising the value of data outputs

To what extent do you think generation and sharing of high quality datasets should be recognised relative to publications?

Please select one option per row

	To a great extent	To some extent	To little extent	Not at all	Don't know
Assessing a researcher's track record in the context of grant applications	?	?	?	?	?
Making decisions in career advancement at institutional level	?	?	?	?	?
Assessing researcher contributions in the context of the REF	?	?	?	?	?

Please briefly explain your ratings.

Please enter your comments in the box below

Priority actions

Please identify up to three priority actions that you think funders and/or institutions should take, if any, to help create a culture that incentivises data access and sharing?

Please enter your answers in the boxes below

1. _____
2. _____
3. _____

Further consultation

As part of the follow on work for this study, we would like to convene a small focus group of researchers and data managers later in the summer to discuss and challenge the emerging findings of this work.

Would you be happy to be contacted to take part in such a focus group?

Please select one option

Yes

No

Please provide your name and email address in the boxes below.

Please note that all responses you have provided will be treated confidentially, and no views you have expressed in the survey will be attributed to you

Name: _____

Email: _____

Complete – Thank you!

STOP

Thank you for completing this survey! If you have provided contact details for further consultation, we will be in touch with you shortly.

Please click the submit button to submit your responses.

ANNEX B: KEY OUTCOMES FROM EXPERT INTERVIEWS

INTRODUCTION

1. The purpose of the in-depth interviews was to explore the key challenges and opportunities around culture and incentives for data sharing with key individuals in research institutions, funding agencies and other relevant stakeholder groups.
2. Interviews were conducted face-to-face or via telephone using a semi-structured approach. A common series of key questions was used as a starting point for the discussions, which were sent to interviewees in advance.
3. Twelve interviews were conducted in total. Interviewees included representatives of HEFCE and JISC, three research deans (or equivalents) in major universities, four post-doctoral researchers, a REF panel chair, a research manager and a senior data manager. The key issues emerging are summarised briefly below.

SUMMARY OF KEY THEMES

4. All interviewees recognised the value and importance of data sharing – with early-career researchers interviewed having particularly strong views in its value in terms of maximising the impact and enabling better science – for example, in terms of the power of studies and quality of analysis.
5. However, there was general consensus that the level of reward and recognition that flows back that flows back to the researcher rarely makes up for the associated costs (both financially and in terms of time). One interviewee asserted that the current situation for data sharing was essentially one where there were “virtually no carrots, and a few sticks – most of them ineffective”. In many cases, sharing data has a significant opportunity cost, and in the absence of recognition or rewards for making data available or sanctions for not doing so, many researchers will understandably not be motivated.
6. It was emphasised that there were important differences between disciplines, and as such it was difficult to generalise. It was highlighted that in a few communities, such as genomics, a culture has developed, whereby for a researcher to be accepted as a member of the “club”, they were expected to make their data available (at least to some extent). But it was felt that this culture did not yet exist across many areas of research. Those interviewed from the engineering and physical sciences suggested that an ethos collaboration and openness was in their experience much more widely developed in their disciplines than in the life or biomedical sciences, where individual researchers tend in general to be more protective and possessive of the data they generate.
7. Interestingly, even in large-scale genomics consortia, there was a feeling that data sharing had limits. One interviewee described how in large scale international collaborations, researchers used their data as “leverage” – indicating that while they were prepared to share aggregate data, they were reluctant to allow access to individual level data and if they did their place in these consortia would be threatened.
8. The Research Excellence Framework was raised by all interviewees as a critical factor in driving the behaviour of universities. It is clear that data outputs have not widely been considered in their own right as part of the REF (and previous Research Assessment Exercise) to this point.

They can in principle be submitted across all panels, but in the 2014 REF the guidance is more explicit in the social sciences that they will be considered than it is in other areas. Because of a widespread feeling (rightly or wrongly) that data outputs would not be viewed as equivalent to publications, there was a view that universities would not encourage their researchers to include them in their submissions. Interestingly, there was a view from one interviewee who had served on a REF panel, that in his area at least, assessors would in fact look favourably on data and other types of outputs and the problem was one of the perception on the part of institutions.

9. For its part, HEFCE has acknowledged that the issue of data outputs will be an important one to revisit in future REF exercises. It is too early to say whether more data outputs will be submitted to the 2014 REF. In considering this issue, a question was raised as to the extent to which a dataset in itself constituted a legitimate and valuable research output. Certainly generation of data on its own is not enough – it must be useable and of use to others. It was felt that one area of the REF where good data management practice at institutional level could (and may already) be recognised is in the research environment component. It was acknowledged however that it would only ever be one of many factors considered here, and that this accounts for only 15% of the REF (compared to 65% for research outputs, and 20% for impact).
10. A further option that has been proposed would be for HEFCE to attempt to mandate data access in submissions to future REF rounds – not only potentially through ensuring that datasets submitted for consideration are accessible, but also that data underlying submitted publications is accessible. Whilst it was acknowledged that this may have a significant impact in terms of changing institutional behaviours, it would be very difficult to implement in practice – with HEFCE interviewees noting that the REF needed to be based on clear and uniform rules, which would not be applicable to datasets where the most appropriate approach for sharing often needs to be considered on a case by case basis.
11. Interviewees acknowledged that the increasing numbers of multi-authors papers being generated from data-rich studies across the biomedical and social sciences did sometimes constitute a concern – particularly in terms of younger researchers receiving recognition for their outputs. On the other hand, the interviews with early career researchers indicated that it was often increasingly necessary and desirable to participate in such consortia to advance their research, and that failure to engage in these activities would perhaps disadvantage them even more. It was also clear that they were judged on their publication record and needed to find ways to ensure they could demonstrate that they had taken a lead on specific research activities. Those interviewed had found ways to achieve this within the consortia in which they worked.
12. This is a challenge that has been faced by researchers in some areas of physics for many years. While interviewees noted that there were working practices in physics to account for this – particularly the inclusion of narrative on CVs and elsewhere to describe a researcher’s contributions – it was still grappling with the issue and had not found a perfect solution. To some extent, large scale collaboration is a fundamental feature of cutting edge research in the particle physics and astronomy fields. Young researchers in these areas are therefore unlikely to have any realistic option but to engage in such consortia, and will be on a relatively level playing field with their peers.
13. At present, physics is the only discipline in the REF with an explicit requirement for researchers to explain their contribution to the papers they submit. Essentially, however, a system exists across all disciplines where panels assess whether a researcher’s contribution to an output has been sufficiently great. If it deems that has, the output is considered as theirs its entirety. If not, it is discounted. HEFCE acknowledged that multi-author papers were likely to become an

increasing feature of other disciplines , and that a similar system to that adopted in physics may be extended to other disciplines moving forward. However, it did not at present feel researchers were disadvantaged by submitting multi-author papers – on the contrary, some of these were the most highly recognised.

14. It was noted that metrics to enable the direct citation of datasets and to track their use were emerging and could have value, but were still in their infancy. It was also noted that there were legitimate questions over the value of some of the new types of metrics being generated. All interviewees felt that awareness of these new types of metrics within the research community was very low.
15. Interviewees noted the key role journals might play in establishing and enforcing good practice in the acknowledgement of datasets. In addition to encouraging deposition of data in recognised repositories where these exist (with an associated requirement for accession numbers), it was also suggested that journals could ensure that, where a dataset is used in course of a piece of research, it is acknowledged and cited properly.
16. Interviewees felt that research institutions had started to recognise the need to support data management and sharing, but the extent to which they actively encouraged and supported it was at present fairly variable. It was noted that universities were beginning to work towards institutional strategies, policies and processes for managing the data generated by research – including development of institutional repositories. However, it was felt that there was still much work to do, and coordination was still lacking in many cases. It was also felt that institutions needed to do more to integrate training in handling and analysing data into undergraduate and postgraduate training.
17. In terms of career advancement and promotion, the feeling was that, data outputs were not currently given serious weighting in decisions. It was suggested that dissemination of data outputs would be viewed positively, but never at the same level as publication output.
18. All of those interviewed from the university sector agreed that career structures for specialist staff needed to support researchers in managing data were severely lacking. Such individuals were usually supported through short-term “soft money” and did not have clear career paths. It was felt that there is a growing consensus that this issue needs attention, but that it had not been prioritised to date. It was also noted that there were few recognised training and professional development opportunities for data managers – although the Digital Curation Centre and JISC provide some visibility and small-scale funding, which has been useful in piloting approaches for training at institutional level.
19. It was also noted that the contribution provided by data managers was often not recognised adequately. The status afforded to data managers and the extent to which they are able to gain authorship credits on publications varies between disciplines and between different groups. It was noted that there was again a possible role for journals here in terms of reaching a consistent position on authorship and contributor description that allowed for data managers’ contributions to be given due recognition.
20. The ethical limits to data sharing were flagged by several interviewees as a key concern (with terms of consent forming a barrier to what it was feasible to share). The need for continued investment in infrastructures and tools to enable safe and secure data sharing was flagged, and the role of Ethics Committees. Interviewees highlighted the need to train Ethical Committees in

this area, and one suggested that Ethics Committees may potentially play an important role in pushing groups to make more data available.

21. In terms of the steps funders could take to encourage data sharing, the key messages were to resource and support it adequately, to explicitly reward those who do it well, and to have clear consequences for those who don't comply with policies. Specific suggestions included:
 - Funders need to ensure they adequately meet the costs associated with managing and sharing data of value that are generated by the research they support – including adequate provision for longer-term preservation;
 - Funders need to develop and sustain the key community data repositories and technical resources (e.g. metadata) needed to enable data sharing;
 - Funders need to establish clear criteria for what data are worth sharing – clear guidance is needed to enable researchers and reviewers to determine which datasets are of real value and worth investing in, and to support researchers in terms of how to deliver in practice.
 - Funders need to be far more effective in monitoring the extent to which researchers and fulfilling their data sharing policies – insisting data are made available and setting in place sanctions (both for where researchers fail to fulfil agreed data sharing plans, and where researchers misuse data generated by others)
 - Funders should develop best practice templates (e.g. for good quality data sharing plans);
 - Funders should recognise and celebrate success – including through case studies, but also perhaps through a high profile prize for data sharing (two interviewees suggested this idea – one describing it as a “data Oscars”)
22. Finally, one interviewee said that in thinking about incentives it was vital to target these effectively. Ultimately, it was argued that group leaders needed to see benefits, as they were the ones with the ability to change practice and bring members of their groups along.

ANNEX C: SUMMARY OF FOCUS GROUP DISCUSSIONS

SUMMARY REPORT

1. Over the course of the project, two focus group discussions were held (each involving 7-8 participants for around two hours) to discuss and challenge the emerging findings of the work:
 - The first focus group held in August 2013 involved a mixed group of seven researchers and data professionals from across the fields of genetics, epidemiology and social sciences;
 - The second focus group held in December 2013 brought together a group of eight postdoctoral researchers across the areas of genetics, epidemiology and cancer research supported by the four partner funders.
2. The format of both focus groups was similar. Participants were given a short introductory presentation to describe the background to the project and the summary findings from the web survey and expert interviews. The session was then opened up for a facilitated discussion – focusing on the key issues addressed in the project.
3. Summary notes of both focus groups follow below.

FOCUS GROUP 1 (AUGUST 2013)

Data management in Institutions

4. There was a strong feeling that widespread deficiencies in data management were symptomatic of a wider problem, in that operational management in research institutions was often ineffective and not sufficiently well prioritised. The situation has been exacerbated by current funding challenges, which have led in many cases to reductions in administrative support.
5. It was suggested that many senior group leaders lack experience of project and budget management, and have in many cases not received any formal training (e.g. PRINCE-2 or other tools). At present, funders do not formally check for project management experience, or assess whether PIs and institutions have effective management structures and skills in place to effectively manage the major research programmes they support, and utilise resources across different teams and studies. It was suggested that institutions could usefully be encouraged to attain recognised international standards (including ISO9001 and ISO2700) to provide assurance that management processes meet recognised quality thresholds.
6. Data management and sharing suffers in this culture. Operational issues and resources are generally viewed as trivial by academics, who tend to focus specifically on generating and analysing data, but not data quality, management and curation more broadly: “professional science and amateur data management”. This is compounded by many smaller studies not being resourced well enough to cater for or support data management, either in terms of funding or access to expertise (including knowledge of repositories, standards, metadata standards, and so forth). PIs tend to be burdened with data management and sharing responsibilities but without much formal support or training. In general, the lack of awareness of quality standards limits the utility of shared datasets – with data often poorly or inconsistently labelled.

7. There was a suggestion that individuals with operational and technical expertise could have recognised positions working alongside PIs to run the operational side of research data management for their teams (together with other key aspects of project management). Their role would include ensuring data are suitably annotated using standardised metadata formats as they are collected, and that datasets are high quality, re-usable, citable and secure.

The role of funders and institutions in planning data management

8. It was suggested that funders could better ensure that researchers provide clear evidence that they will have access to the necessary infrastructure and skills required to manage and share their data, when considering data management and sharing plans set out in applications. This would in turn provide a clear incentive for institutions to improve their support and infrastructure for data management.
9. There was clear agreement that funders needed to play a much more active role in monitoring and providing support for data management and sharing throughout the project. A model was suggested where:
 - once a project has been selected for funding, funders would discuss and negotiate the data management plan with researchers – identifying the most suitable repository to host the data, and clearly agreeing which data will be made open and which will not;
 - funds would then be clearly earmarked for the data dissemination based on the agreed approach; and
 - at the end of the project, the database could be submitted to the funder - who could potentially have it reviewed by a dedicated panel and place it in the agreed repository (after an agreed grace period where appropriate to allow time for first publication).
10. There was agreement within the group such a system would be effective in funders ensuring that that data generated is high quality, re-usable and citable. However, to implement such a system, and play an active role in data assurance post-award, funders would need to take on a completely new function. The capacity and skills to fulfil this are not currently in place, nor are they currently conducted by existing repository infrastructure, such as the UK Data Archive, for research areas where these services exist.
11. This also raises broader questions of where ultimate responsibility (and indeed ownership) for data lies between funders, institutions and individual researchers, and such a process would also pose issues around intellectual property and commercial use. The relationship between these stakeholders is complex, and it is not always clear where responsibilities should lie. Nevertheless, there was felt to be a clear onus on funders to do more and be clearer on their expectations, rather than just leave this to institutions.
12. It was acknowledged that funding decisions are made by committees of experts, based on peer review – most of the individuals involved will lack the knowledge to adequately assess data management issues. The Group agreed with the finding of the survey that data management plans were often viewed as superficial and not taken seriously. There was a recommendation, popular within the group, that a technical secretary could be appointed to each peer review panel, to provide expert input on data management issues. Such an approach would also provide a voice for data managers and could help to enhance their status.

Recognition and reward

13. There was agreement that funders should give explicit acknowledgement when making funding decisions to an applicant's track record of sharing useable datasets. Investigators working in data intensive fields could be encouraged to be early adopters when they serve as reviewers.
14. It was suggested that HEFCE could usefully make a strong statement to researchers as soon as possible of plans to make data sharing a clear component in the post-2014 Research Excellence Framework, even if the details have yet to be fully worked out. Funders should collectively advocate in favour of HEFCE adopting this approach and support them in doing so. A clear move of this type would make a major difference in pushing researchers and institutions to ensure the resources and skills were in place in order to build their track records in this area.
15. There was a view that the emergence of data journals could help to enable well-described, consistently annotated datasets to be recognised as research outputs, alongside publications. Providing data managers are given lead authorship status for such papers, they may also provide a more effective means of assigning credit. These papers would also help to enable the use of data to be cited and tracked.
16. It was noted that promoting wider awareness and uptake of other systems and metrics to track the use of datasets (such as DataCite DOIs) was a difficult challenge – and a 'chicken and egg' situation, in that researchers are unlikely to spend the time using them in the absence of formal recognition, whilst funders and institutions won't provide the recognition in the absence of better and more widely-used metrics.
17. There was a clear view that young post-doctoral researchers do face difficulty working in fields where large multi-author papers are more common. Where they exist, current systems for assigning contributor roles were felt to be largely uninformative. Whereas more advanced systems do exist in particle physics and astronomy, it was noted that these were far from perfect – and that challenges remain in these fields, particularly for the 'long-tail' of research outside the major large-scale centres.

Careers

18. Participants agreed strongly with the finding of the research that there is an absence of clear career structures for data managers. There was a suggestion that funders could establish a dedicated fellowship scheme for leaders in data management – this would need to be flexible in order to accommodate those coming from academic pathways and those who come from an IT, software or project management. Some form of professional qualification and accreditation may also have value. There was recognition that data managers were a diverse community, from a variety of different types of background and it was emphasised that their motivations may often differ from those of academics – i.e. being named on publications may not be important to many. A clearer career structure and role – with opportunities for promotion and adequate remuneration for good performance – would however be welcomed by most.
19. It was noted that early career researchers are not well informed about the issues associated with data management and sharing. As seminars are often held in institutions, which are frequently attended by junior researchers, one idea was to use seminars on these issues to help raise awareness.

Infrastructure

20. It was recommended, and widely agreed, that the establishment of common repository infrastructure (“a PubMed for data”) would be very useful. For a model see HASSET (the Humanities and Social Science Electronic Thesaurus developed by UKDA), which facilitates searches for data by listing controlled vocabularies and terminology for variables. Such a resource would:
 - enable researchers to search for links to data in which they are interested
 - increase citability and discoverability
 - facilitate meta-analyses of datasets
21. It was noted that the European Commission has developed an online portal for open access papers, and is planning to add databases to this portal – it was suggested that Wellcome and other funders could pair up with them for this project.
22. There are potential economies of scale to be had by institutions in sharing infrastructure, if the benefits could be made clear, and not perceived as a threat to competitiveness. It was emphasized however that other drivers are pushing institutions into greater competition.

Creating an enabling environment

23. Part of the difficulty in creating incentives lies in uncertainty over the market for data sharing: potential users need to be on board at the outset, with an understanding of the potential for exploitation with existing datasets that have not been thoroughly interrogated. Promotion, discoverability and financial support for secondary science will be crucial.
24. It was noted that technical knowledge of data sharing among ethics committees was currently extremely low. An interesting question for the future will be the stance that these committees take on data sharing – in terms of the balance between safeguarding privacy and security while enabling data to be used for potentially beneficial secondary uses. If they lack understanding they may tend toward caution and conservatism.
25. It was widely acknowledged that there is a need for a quasi-public debate on information use, sharing and risk from shared datasets. Good data quality and security should enable datasets to be shared safely, but a dialogue about the scope and benefits of sharing would be in the public interest. There also needs to be dialogue on the legal issues and liability surrounding the sharing of sensitive data and risks of re-identification, involving researchers, institutions, participants, ethics committees and data repositories.

FOCUS GROUP 2 (DECEMBER 2014)

The value of data sharing

1. With regard to the value of data sharing, there was strong agreement that it was important to encourage as much data sharing as possible, but that the value can be hard to anticipate in a particular case.
2. Researchers need to make pragmatic and realistic decisions about which data are worth sharing. At the moment, it is problematic to determine likely user demand for data. Participants noted that in many cases there are no resources to help users look for data that might be of interest. Data discoverability was a key challenge.
3. It was emphasised that data sharing was only of value if the data were useable and provision of sufficient metadata was key. Again decisions were required as to what level of metadata was appropriate for a particular dataset, in light of its value.

Infrastructure to support data sharing

4. With regard to data infrastructures, it was noted that in some cases central repositories could be used and in others bespoke databases would be required. Funding to maintain and develop systems would be essential in both cases, and it was noted that some existing infrastructures were under-resourced and difficult to use (the EGA was mentioned as one particular example). In exploring models for developing and supporting the required infrastructure, participants suggested a potential role for commercial suppliers and service-providers that should be explored (recognising that issues around sustainability would need to be explored).
5. In many areas, it was noted that researchers may lack existing knowledge and support of how to make data available. There may be potential for funders to do more to direct researchers to available resources and potentially provide additional sources of advice.
6. It was noted that there is an increasing need for academics and postdocs to develop skills in managing and analysing data. There was also a need for specialist skills at institutions that sits above group level, and funders potentially need new approaches to support this.

Career development and recognition

7. It was emphasised that data sharing was hard to prioritise for junior researchers given that they were assessed primarily on their publication record – finding the time was difficult.
8. There was agreement that, while involvement in large-scale collaborative projects likely to generate multi-author papers was an increasing feature of genomics and epidemiology research and where much of the best science took place, early-career researchers were still expected to generate first and last author papers to progress. Indeed, some funders specifically asked for papers of this type to be listed separately. Therefore, early-career researchers needed to ensure that they had opportunities to achieve this balance.

9. There was overall consensus was agreement that reliance on author position needed to be assessed, and funders needed to move away from this. It was noted that recognising contributions to multi-author papers didn't need to be an issue – with some disciplines, notably particle physics, having developed ways around this which worked. It was also noted that decisions on career advancement might be more sophisticated in recognising contributions to large-scale initiatives than fears justified.
10. It was emphasised that senior academics had a responsibility to ensure that young researchers in their teams were given appropriate vocational care and didn't unduly bear the burden of data sharing requirement in a manner which effected their career prospects. It was noted that changing incentive structures would have little effect on younger researchers if they fundamentally didn't want to be spending their time making data available.
11. Further to this point, it was noted that many group leaders had a tendency to try to hire as many postdoctoral staff as they could within a limited budget, with the result that data sharing was often passed onto them when it wasn't what they had anticipated or wanted. It was argued that greater honesty was needed about the roles required (in particular, that if specialist data managers were needed rather than postdoctoral researchers that this was requested). Funders needed to ensure that their support enabled specialist data managers to be appointed if required, and that staff requested were appropriate.
12. There was recognition that different metrics (other than publications) were probably needed to assess the performance and contributions of data managers, rather than trying to focus on ways that these individuals could receive credit through these systems.
13. There was discussion over possible mechanisms to enable the citation of datasets – ideas included a separate section on papers (or on PubMed) to reference any data generated by others underpinning the research.
14. There was general agreement that there needed to be some form of additional recognition for setting up a large-scale dataset which was used by others. There were particular issues on studies undertaken in low- and middle-income countries in terms of capacity building and ensuring that researchers based in these settings had due opportunities to use and gain credit for datasets they had put the resource into generating.