

ESTABLISHING INCENTIVES AND CHANGING CULTURES TO SUPPORT DATA ACCESS

May 2014

ACKNOWLEDGEMENT

This is a report of the Expert Advisory Group on Data Access (EAGDA). EAGDA was established by the MRC, ESRC, Cancer Research UK and the Wellcome Trust in 2012 to provide strategic advice on emerging scientific, ethical and legal issues in relation to data access for cohort and longitudinal studies.

The report is based on work undertaken by the EAGDA secretariat at the Group's request. The contributions of David Carr, Natalie Banner, Grace Gottlieb, Joanna Scott and Katherine Littler at the Wellcome Trust are gratefully acknowledged. EAGDA would also like to thank the representatives of the MRC, ESRC and Cancer Research UK for their support and input throughout the project. Most importantly, EAGDA owes a considerable debt of gratitude to the many individuals from the research community who contributed to this study through feeding in their expert views via surveys, interviews and focus groups.

The Expert Advisory Group on Data Access

Martin Bobrow (Chair)
James Banks
Paul Burton
George Davey Smith
Rosalind Eeles
Paul Flicek
Mark Guyer
Tim Hubbard

Bartha Maria Knoppers
Mark McCarthy
Andrew Morris
Onora O'Neill
Nigel Shadbolt
Chris Skinner
Melanie Wright

EXECUTIVE SUMMARY

This project was developed as a key component of the workplan of the Expert Advisory Group on Data Access (EAGDA).

EAGDA wished to understand the factors that help and hinder individual researchers in making their data (both published and unpublished) available to other researchers, and to examine the potential need for new types of incentives to enable data access and sharing. This is a critical challenge in achieving the shared policy commitment of the four EAGDA funders to maximise the benefit derived from data outputs and the considerable investment they have made over recent years in supporting data sharing.

In addition to a review of previous reports and other initiatives in this area, the work involved in-depth interviews with key stakeholders; two focus group discussions; and a web survey to which 35 responses were received from a broad range of researchers and data managers.

Although based on a relatively modest number of responses and interviews, the findings closely mirrored those of previous work in this area. In particular there was a clear, overarching view that the research culture and environment is not perceived as providing sufficient support, nor adequate rewards for researchers who generate and share high-quality datasets.

Our findings were that:

- making data accessible to others can carry a significant cost to researchers (both in terms of financial resource and the time it requires) and there are constraints in terms of protecting the privacy and confidentiality of research participants;
- while funders have done much valuable work to encourage data access and have made significant investments to support key data resources (such as the UK Data Service for the social sciences), the data management and sharing plans they request of researchers are often not reviewed nor resourced adequately, and the delivery of these plans neither routinely monitored nor enforced;
- there is typically very little, if any, formal recognition for data outputs in key assessment processes – including in funding decisions, academic promotion, and in the UK Research Excellence Framework;
- data managers have an increasingly vital role as members of research teams, but are often afforded a low status and few career progression opportunities;
- working in data intensive research areas can create potential challenges for early career researchers in developing careers in these fields;
- the infrastructures needed to support researchers in data management and sharing, and to ensure the long-term preservation and curation of data, are often lacking (both at an institutional and a community level).

The need to place a greater value on data outputs has been recognised in a succession of reports and discussions on enabling data sharing stretching back well over ten years. Yet, despite this, very few concrete steps have yet been taken to achieve this.

Our goal in this work was not to restate the challenges, but to identify ways forward. While this is a difficult problem, if we are to achieve the goal of maximising the value of research data, it simply cannot be ignored. We believe that there are specific actions that research funders and research leaders can progress now through which they could make a tangible impact in working with the wider community to drive the fundamental change in culture that is required.

At this stage, we believe that the primary focus should be on developing mechanisms that encourage and reward good practice, rather than on penalising researchers who fail to fulfil their planned approaches for sharing data. It is premature to introduce punitive sanctions when the infrastructures and tools available to researchers may act as a constraint, and where adequate resources and costs may not have been provided. This situation should be kept under review, and sanctions may well have a role in the future.

While our recommendations are primarily addressed to the UK community, this is clearly an issue that is international in scope. It is vital therefore that funders and research leaders maintain an active on-going dialogue with international partners and work with them to build common incentive structures and effect cultural change.

RECOMMENDATIONS

We recommend that **research funders** should:

1. Strengthen approaches for scrutinising data management and sharing plans associated with their funded research – ensuring that these are resourced appropriately and implemented in a manner that maximises the long-term value of key data outputs.
2. Urge the UK Higher Education funding councils to adopt a clear policy at the earliest possible stage for high quality datasets that are shared with others to be explicitly recognised and assessed as valued research outputs in the post-2014 Research Excellence Framework
3. Take a proactive lead in recognising the contribution of those who generate and share high quality datasets, including as a formal criterion for assessing the track record and achievements of researchers during funding decisions.
4. Work in partnership with research institutions and other stakeholders to establish career paths for data managers.
5. Ensure key data repositories serving the data community have adequate funding to meet the long-term costs of data preservation, and develop user-friendly services that reduce the burden on researchers as far as possible.

We recommend that **research leaders** should:

6. Adopt robust approaches for planning and costing data management and sharing plans when submitting funding applications.
7. Ensure that the contributions of both early-career researchers and data managers are recognised and valued appropriately, and that the career development of individuals in both roles is nurtured.
8. Develop and adopt approaches that accelerate timely and appropriate access to key research datasets.
9. Champion greater recognition of data outputs in the assessment processes to which they contribute.

We also emphasise that research institutions and journals have critical roles in supporting the cultural change required.

Specifically, we call for **research institutions** to develop clear policies on data sharing and preservation; to provide training and support for researchers to manage data effectively; to strengthen career pathways for data managers; and to recognise data outputs in performance reviews.

We call on **journals** to establish clear policies on data sharing and processes to enable the contribution of individual authors on the publication to be assessed, and to require the appropriate citation and acknowledgement of datasets used in the course of a piece of published research. In addition, journals should require that datasets underlying published papers are accessible, including through direct links in papers wherever possible.

INTRODUCTION

AIMS AND OBJECTIVES

1. This project set out to examine the extent to which the costs to researchers and research teams of providing access to datasets are balanced by the benefits (in terms of recognition, rewards and career advancement) that they receive. The goal was to identify possible areas in which additional incentives might be required, in order to foster a culture that supports researchers in making high quality data available and gives due recognition to those who do it well.
2. The project had three primary aims:
 - to assess the costs and benefits associated with making data available to others, and the current factors that may act as disincentives for researchers in the field – both in sharing data and in developing careers in data-intensive fields;
 - to review the current status of established and emerging mechanisms to incentivise timely and appropriate sharing of data;
 - to develop recommendations and best practice guidance for EAGDA funders and studies on how they can establish appropriate incentives to maximise access and use of data.

METHODOLOGY

3. The initial research phase of the project involved three main components.
 - First, a **literature review** was undertaken to bring together information from recent reports and on-going initiatives which have addressed these issues.
 - Second, a **web survey** was designed to gather the perspectives of researchers and data managers on the current barriers to data sharing and the areas in which incentives may be required. The survey was sent to a sample of 110 researchers and data managers identified by the four partner funders in late June 2013. Thirty-five responses were received by the closing date of 19 July 2013 – with respondents spread fairly evenly across the areas of genetics, epidemiology and social sciences. A full summary of the results of the survey can be found in **Annex A**.
 - Third, twelve in-depth **informant interviews** were conducted with several key stakeholders. Those interviewed included representatives of two funders (HEFCE and JISC), four senior academic managers at major universities, four postdoctoral researchers, a chair of a REF panel and a senior data manager. A summary of the issues raised in the interviews can be found in **Annex B**.
4. Based on the outputs of this research, an initial series of conclusions and possible options were identified. Two small **focus group discussions** were held – which brought together key stakeholders to discuss and challenge the emerging findings of the work.

The first focus group (held in August 2013) involved a varied group of researchers and data managers. The second (held in December 2013) involved a group of early-career researchers in the fields covered by EAGDA. Summary reports of the two focus group discussions can be found in **Annex C**.

5. The outcomes of the focus groups were used to further refine the draft findings and recommendations, which were shaped through on-going discussions with EAGDA members through the project. The final recommendations of the work were discussed and agreed by EAGDA at its March 2014 meeting.
6. This report summarises the key findings and conclusions of this work. It begins by describing the challenges around creating an enabling culture for data access and sharing that have been identified in previous discussions and research on this topic, and the different areas in which it has been suggested that incentives and rewards may be required. It then discusses the evidence gathered through the research undertaken for this project, before outlining EAGDA's conclusions and recommendations.

A NOTE ON TERMS

7. Throughout this report, the term 'data generator' is used to refer to those who have undertaken original research studies that have resulted in data resources that are of potential value to others. The term 'data user' describes those who wish to access and use data resources generated by others, whether for research or other purposes. It is recognised of course that many researchers are both data generators and data users.

CULTURE AND INCENTIVES FOR DATA ACCESS AND SHARING: THE ISSUES

DATA ACCESS: OPPORTUNITIES AND COSTS

8. Over recent years, issues surrounding the management, preservation and sharing of research data have been the subject of rigorous policy debate. Increasingly vast and complex datasets being generated across the scientific enterprise are creating profound challenges for data storage, access and interpretation. At the same time, there is a growing recognition that these rich datasets hold enormous untapped potential, and that enabling them to be more widely accessed and used could generate valuable new insights and discoveries.
9. The Royal Society's *Science as an open enterprise* report, published in June 2012, discussed how these trends were transforming the research enterprise¹. The report called for a fundamental shift away from a culture where data are regularly seen by researchers as a 'private preserve'. It highlighted the tremendous opportunities in a new era of data-driven science, which it argued needed to be underpinned by a culture of 'intelligent openness' in which datasets generated by research are accessible, intelligible, assessable and useable.
10. The MRC, ESRC, Cancer Research UK and the Wellcome Trust have long recognised this potential, and have had policies in place for several years that require data outputs of value to the broader research community to be made available as widely as feasible, so that they can be accessed and used to maximise the resulting public benefit. Their policies closely mirror those of other major research funders around the world, and there is a growing international consensus on the importance of increasing the availability of research datasets².
11. It is clear that the success of funders in delivering these policies will depend upon the creation of an enabling environment. Key elements of this are that researchers have access to the resources they require to make data available, and that they operate within a culture which balances the interests of data generators, data users and research funders. It has long been recognised that many actors have critical roles and responsibilities in fostering this environment^{3,4,5}, and that different disciplines are at different stages in terms of the level of infrastructure and support available.
12. A critical feature of this enabling environment is appropriate incentives which ensure that those who put the time and effort into creating high quality data sets and making them available to others receive due recognition and reward for their efforts, and are not disadvantaged relative to their peers. Policy discussions on this topic over several years have consistently found that the lack of adequate incentives is a major limiting factor constraining efforts to increase access to research data^{1,3,6}.

“...major barriers to widespread adoption of the principles of open data lie in the systems of reward, esteem and promotion in universities and institutes. It is crucial that the generation of important datasets, their curation and open and effective communication is recognised, cited and rewarded.”

*from Science as an open enterprise
(Royal Society, June 2012)*

13. In examining this topic, it is important to consider where the benefits and costs of sharing data lie. The benefits of increasing access to research datasets at a community level are well-articulated and include:
 - allowing the evidence underlying published scientific findings to be more readily verified and scrutinised;
 - reducing duplication of effort;
 - enabling secondary uses of data – including in ways that might never have been anticipated by the original data generator.
14. Arguably, however, the benefits at the level of the individual data generator might be somewhat less clear cut. There is a valid argument that it can raise the visibility of their work (in some cases, for example, leading to new collaborations) and would be recognised as a valuable service to the community by their peers. There is also evidence that sharing data underlying published findings generates increased levels of citation for the work⁷. But these benefits could seem less tangible, and might in themselves be unlikely to drive many researchers to place additional resource into making their data more widely available.
15. On the other hand, there are a number of very real costs and risks for data generators associated with making data available, in particular:
 - getting data into a format in which it can be used effectively by others may carry significant resource implications – both in terms of financial cost and staff time. For example, in a major survey of over 1,300 scientists in 2010, 53% of respondents who answered the question cited lack of time as their main reason for not sharing data, and 40% lack of funding⁸. Researchers may face a choice as to whether they divert resources from their research programmes to undertake data sharing.
 - researchers who share data run the risk that they will sacrifice some of the intellectual credit that they could have otherwise gained from the use of the data⁹ – particularly if those who access the data have similar research interests and greater capacity for data analysis. This may be a particular concern where users based in well-resourced institutions seek to access data generated by researchers in resource-poor settings.
 - researchers may face a reputational risk if data are found to be inaccurate, are used in an inappropriate or unethical manner, or are misinterpreted.

- researchers may have legitimate concerns that sharing data may risk the privacy and confidentiality of research participants, or not be consistent with the consent they have given, and tend toward a cautious approach.

16. Therefore, without additional incentives, researchers may often be disinclined to make data sharing a priority. A number of different areas have been identified previously in which incentives could play a role. These formed the basis for this research, and are discussed briefly below.

MEETING THE FINANCIAL COSTS OF DATA MANAGEMENT AND SHARING

17. An increasing number of funders (including the four EAGDA funders) now require researchers to include a data management and sharing plan within grant applications. Data plans are expected to describe how and when data of wider value that results from the research will be made available. Funders commit to review these plans as an integral part of the funding decision; and, if the grant is successful, to meet the costs associated with the agreed plan within the overall award.

18. However, the extent to which this approach is effective in practice is open to question¹⁰. Certainly, concerns remain that funders are not adequately meeting the costs incurred by researchers in managing their data and enabling wider access.

19. The financial costs associated with managing and sharing data will of course vary considerably depending on the nature of data generated and the approach for preserving and sharing data that is needed to maximise its value. Quantifying these costs on a generic basis is challenging, however the Keeping Research Data Safe project provides one valuable model¹¹. This work sought to evaluate and model the costs and benefits (both short and long-term) of preserving social science data. One major conclusion was that, for data repositories, ingest and acquisition constitute a significantly higher proportion of the overall cost than long-term preservation.

PROTECTING INTELLECTUAL OUTPUTS FOR DATA GENERATORS

20. Most funder policies recognise that data generators will normally have the right to a reasonable (but not unlimited) period of exclusive use of the data. For many funders (including MRC, Cancer Research UK and Wellcome Trust), this is considered on a case-by-case basis depending on the nature of the data and of the research. However, some do set specific timeframes: the ESRC, for example, requires that researchers must offer their data to the UK Data Service within three months of the end of the project.

21. Alongside this notion of 'protected periods' of exclusive access, data generators may agree to share their data with users, but place specific conditions on its use. In some situations for example, users may only be able to access the data in the context of a collaboration with the data generator. Data access agreements typically require that the data resource be acknowledged in publications, and many stipulate that outputs (including copies of publications and derived data) are returned by the user.

22. In the genomics field, a culture of making datasets that constitute community resources rapidly available to users has become quite well embedded¹². For major genomic studies, the concept of a protected period for data generators - over which they retain the right to publish their primary analysis of the dataset is well established¹³. This protected period is commonly known as a 'publication moratorium'. For some studies, a specific maximum time period (typically nine or 12 months) is defined. For others, the moratorium period can be open-ended. In March 2014, the ENCODE initiative updated its data policy to remove the moratorium period and allow users to publish results based on the data immediately²⁰. The policy encourages researchers using unpublished ENCODE data to contact the data producers to discuss coordinated publications, but makes clear that this is optional.

MEASURING AND RECOGNISING THE USE OF DATA OUTPUTS

23. There are a number of emerging mechanisms that may enable data generators to more effectively monitor the downstream use of their data^{14,15,16}:
- An increasing number of specialist data journals now exist in which researchers can publish papers providing descriptive detail on their datasets, which can be cited in the normal manner by data users. Examples include Scientific Data (published by Nature Publishing Group) and the Journal of Open Public Data Health (published by Ubiquity).
 - Tools that enable the citation of datasets themselves through the attribution of Digital Object Identifiers (DOIs) and other permanent identifiers for data are gaining increasing traction. Systems to provide persistent identifiers for researchers - such as ORCID (the Open Researcher and Contributor ID) – are also becoming increasingly widely adopted. Linking ORCID identifiers to DOIs and other identifiers holds the potential to enable researchers to gain a more complete picture of their outputs.
 - New types of impact metrics allow the collation of information on a broader range of research outputs, including data - of particular note in the context of this discussion is the Data Citation Index established by Thompson Reuters in late 2012, and the Bio-Resource Impact Factor developed by the GEN2PHEN programme¹⁷.
24. It is important to emphasise that this is still a relatively young area and while several initiatives have made impressive progress, uptake of these systems is not yet widespread in many research areas. Furthermore, to act as an incentive, such metrics need to be given explicit recognition in research assessment processes (in the context of funding, career advancement and the Research Excellence Framework). At present, publications dominate other types of outputs in terms of the weighting they are given.
25. Even when considering the assessment of publication outputs, there are concerns that researchers undertaking large scale and longer-term studies involving generation of data resources may sometimes be disadvantaged. These studies can result in published outputs having large numbers of contributing authors. There are concerns that this may potentially act as a disincentive for early-career researchers in some fields – particularly

those that typically use first, second or last author position as a proxy to judge whether a researcher has taken a leading role in the work.

BUILDING KEY SKILLS AND RESOURCES TO ENABLE DATA SHARING

26. In addition to current assessment methods providing potential disincentives for researchers, it is also widely acknowledged that they provide few incentives for data managers – who may make vital contributions to research teams in terms of managing, curating and analysing datasets, but are often not given authorship credits or other forms of recognition. Furthermore, reports over several years have highlighted the lack of defined career paths for data managers¹⁸.
27. In addition to access to skills, researchers and their teams need to be able to access the tools and resources they need to share data effectively¹⁹. Some research fields are well served by existing community data repositories – including those provided by the UK Data Service for social sciences and the European Bioinformatics Institute for genomic and biomolecular data. Both of these resources provide unified points of access for researchers and other users to high quality datasets, and additional services (such as the Secure Lab of the UK Data Service to enable access to data that are particularly detailed, sensitive or confidential). In addition to providing for the long-term preservation and curation of datasets of value, community repositories also play an important role in developing and rolling out appropriate data standards, metadata formats and tools. Previous surveys have indicated that for many fields, a lack of recognised data standards and limited awareness of standards where they exist may form a significant barrier to data sharing⁸.

IMPLEMENTING PENALTIES AND SANCTIONS

28. It has been suggested that the introduction of clear sanctions for non-compliance with funder and institutional policies on data sharing would help to ensure researchers fulfil policy requirements¹⁸. Indeed, the ESRC's policy is already to withhold the final payment on a grant, if the grant holder fails to offer their data to the UK Data Service for archiving within three months of the end of the project. It is therefore important that any discussion of incentives also considers the role of punitive mechanisms alongside reward-based approaches.

KEY FINDINGS

29. The key findings emerging from the survey, interviews and focus group discussion are summarised briefly below, under the following headings:

- Data sharing – the value, the cost and the barriers
- The role of funders
- The role of research institutions and the Research Excellence Framework
- Careers issues for data managers
- Career development issues for postdoctoral researchers
- Role of publication moratoria
- Data citation and metrics
- Other issues identified

DATA SHARING – THE VALUE, THE COST AND THE BARRIERS

30. The survey and interviews provided a strong indication that researchers and other stakeholders across the fields supported by the EAGDA funders recognise the importance to the research enterprise of ensuring key data resources are made widely available (**Box 1**).

Box 1 – the importance of sharing data (illustrative quotes from survey respondents)

“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.”

“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.”

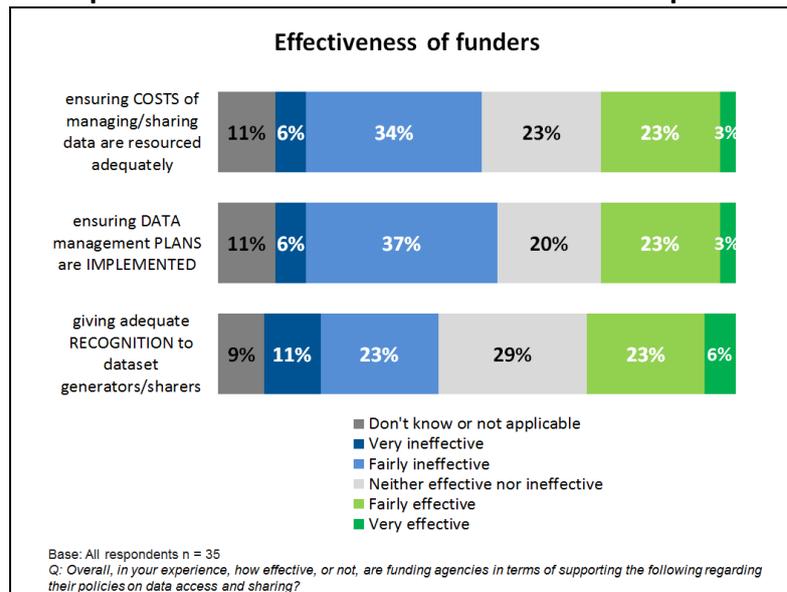
31. However, they also emphasised that making data more widely available in many cases carries a significant cost, and that major barriers exist. In the web survey, around two thirds of respondents (66%) indicated lack of time as a constraint. In addition, a majority of respondents agreed that protecting the privacy of research participants (60%); lack of recognition (55%) and shortage of funds (51%) also formed barriers.

32. Limitations on technical resources (including standards and databases) were identified by 49 per cent of respondents as a barrier, and the need for resource infrastructure and more user-friendly tools was a theme highlighted in both focus group discussions. The idea of a common overarching repository infrastructure for data was proposed.

THE ROLE OF FUNDERS

33. There was a perception from many survey respondents that data management and sharing plans are not being assessed adequately in funding decisions – both in terms of their content and their cost. Many respondents also held the view that funders did little to monitor or enforce the implementation of data sharing plans post-award and ensure that researchers were delivering these in practice (**Graph 1** and **Box 2**).

Graph 1 – views on the effectiveness of funder policies



Box 2 – funder policies

“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.”

“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.”

34. These views were strongly endorsed by the first focus group, at which participants suggested a potential model where funders engaged in a much more active negotiation with grant recipients pre-award to agree and cost data management and sharing plans. It was proposed that at the end of a grant, funders should take direct responsibility for reviewing the resulting dataset and ensuring it was deposited in the appropriate repository. It was also suggested that funders could appoint dedicated experts on data

management issues to funding committees to assist in the technical review of data sharing plans, given the perception that committees often lacked this specific expertise and would not prioritise this.

35. Survey respondents, interviewees and focus group participants also suggested that funders needed to provide greater, and more sustained support for key data repositories and work with them to develop more user-friendly services for researchers. There was also felt to be a related need for funders to provide more specific guidance and assistance to researchers – for example, in terms of shared resources and expertise they could draw on to assist in managing and sharing data, and in determining which data had value.
36. There was a widespread view among those consulted as part of this work that previous track record of sharing data was not given any serious level of recognition in funding decisions. There was also a suggestion that funders could potentially do much more over and above this to recognise those who generate and share datasets of value. This might include highlighting success stories and recognising achievement formally or developing a dedicated award or prize for data sharing.

THE ROLE OF INSTITUTIONS AND THE RESEARCH EXCELLENCE FRAMEWORK

37. There was a general consensus that many institutions are beginning to recognise the importance of data management and sharing, and making some progress in developing institutional policies and resources (**Box 3**). However, the overall impression was that there was still some way to go. Interestingly, many interviewees and survey respondents were either not aware whether their institution had a data sharing policy, or thought that it did but weren't familiar with the content – indicating perhaps that institutions have more work to do in communicating their policies to their staff.

Box 3 – Institutional support for data management and sharing

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

“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)”

38. Many survey respondents indicated that for the most part they did not feel that their institution encouraged or supported them in making their data available, or in planning their approach when preparing funding applications. There was also felt to be a need for training at institutional level from undergraduate level upwards to equip researchers

to manage data, together with provision of technical assistance to use data resources – particularly for smaller research groups.

39. There was a detailed discussion at the first focus group around institutional culture. Participants expressed a strong view that the lack of support for data sharing was symptomatic of a wider lack of formal operational and project management skills in many research groups. In the context of data management, this led to a focus on data generation, with little attention given to preservation and curation. This ethos, combined with the lack of training for researchers, was felt to have contributed to a continued lack of awareness and focus on the importance of metadata, which has limited the utility of many shared datasets.
40. As in funding decisions, there was a widespread view that data outputs are not given any serious value in the context of career advancement decisions in institutions. Only 11 per cent of survey respondents felt data was recognised to some extent, relative to publications.
41. It was emphasised that the Research Excellence Framework (REF) is a critical factor in driving incentive cultures in universities. At the current time, whilst data inputs may be submitted for consideration in the REF, this was felt not to happen to any great extent in practice. Some felt that there may also be a perception (rightly or wrongly) in institutions that such inputs would not be looked upon favourably, which might result in them discouraging researchers from submitting them.
42. It was suggested by several of those consulted that more formal recognition for data outputs in the REF could go some way towards changing institutional behaviour. A mandate in the REF for open data would almost certainly shift institutional behaviour but there was some caution that this would be problematic to implement in practice.

CAREERS ISSUES FOR DATA MANAGERS

43. There was a strong message that while data managers have a vital role to play within research teams for data-intensive studies, they do not currently have clear career paths in institutions or access to training and development opportunities. In addition, it was felt that their contributions were often not adequately recognised (for example, in terms of being included as authors on studies). While it was emphasised that the situation varied between different research groups, there was an overarching message that data managers are often not given the status their roles or contributions warrant (**Box 4**).

Box 4 – Career issues for data managers

“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 are anonymised before issue.”

44. The impression from the interviews conducted with senior staff at universities was that there was a general recognition of the need to nurture these skills, but that few had yet taken concrete actions to address these issues. Several of those consulted indicated that data managers were often dependent on short to medium term grant funding, rather than being supported at institutional level across projects.
45. In the focus group discussions it was emphasised that data managers should not be considered as a homogeneous group. Data managers come from a variety of backgrounds – in particular, some come into the role from academic pathways and others from IT, software or project management paths. Data managers may have quite different motivations and drivers to academic researchers, and standard ways of recognising contributions (i.e. authorship on publications) may be less important. There was also a strong feeling more generally that the performance and contribution of data managers should be based on different metrics to those of academic researchers.

CAREER DEVELOPMENT ISSUES FOR POSTDOCTORAL RESEARCHERS

46. In the interviews and focus group discussion with early-career researchers, it was acknowledged that the increasing trend toward large multi-author publications in data-intensive fields could sometimes form a constraint in terms of making it harder to demonstrate a leading contribution. At the same time, however, participation in high-profile large-scale initiatives in the genomics and epidemiology fields was often felt to be essential – with several indicating that it was where the best science was taking place. Indeed, in some areas (especially genomics), it was felt it would be seen as a distinct disadvantage to young researchers not to have made a contribution to these major partnerships.
47. There was a consensus overall that there was a pressure on early-career researchers to balance participation in these initiatives with a continued need to produce papers in which they could demonstrate they had played a leading role (and which was still largely assessed through first and last author status). Many of those consulted indicated that they had found ways to successfully achieve this balance. It should be noted, however, that there could be a degree of self-selection here - in that those engaged were all

current postdoctoral researchers, and we did not attempt to identify individuals who might have been less successful - for example, those who had left the field.

48. Interviewees and focus group participants noted that there are potential lessons that could be learned from other disciplines, particularly particle physics and astronomy, which have decades of experience of dealing with the issue of large-scale collaborations and multi-author papers. Systems exist in these disciplines, and are reflected in the REF and other assessment processes, in which researchers provide some narrative to describe their individual contributions to papers. There was a view however that there is no perfect solution even in these areas.
49. Some survey respondents supported further development of systems to tag the contributions of individual authors on publications. In any event, there was a strong view that funders and researchers needed to move away from a perceived reliance on using first and last author status as a proxy indicator of a leading research contribution – a practice which, it was suggested, is becoming increasingly outdated.
50. A further key message from the focus group with early-career researchers was that there was a risk that the burden of data management could unduly fall on postdoctoral researchers. There was a perception that study leaders were motivated to keep costs low, and tended to hire the maximum number of postdoctoral staff they could within this – often giving these individuals responsibility for meeting data sharing obligations, when it wasn't what they wanted to do, and it didn't advance their careers as researchers.
51. It was felt there was a need for more honesty on the different roles required, and for group leaders to seek funding for specialist data managers where they were required. More generally there was a feeling that group leaders had a core responsibility to both postdoctoral researchers and data managers to nurture their career development and ensure they were given the opportunities to develop in their roles.

ROLE OF PUBLICATION MORATORIA

52. Publication moratoria were felt in general to be a fairly effective system in enabling sharing of datasets, whilst protecting the data producer's right to publish first. The overarching view was that there may be scope to trial the use of these systems in other fields where they are not commonly used. It is important to note, however, that a small minority of survey respondents flagged strong concerns relating to the potential misuse of moratoria, specifically in terms of better-resourced groups using them in a way that unfairly disadvantaged smaller groups.

DATA CITATION AND METRICS

53. The evidence from the survey responses was that awareness of emerging identifier, citation and metric systems among the communities served by EAGDA was extremely low. For example, less than half of respondents (46%) indicated that they were aware of

DataCite DOIs, and only 11 per cent had actually used them. The interviews and both focus group discussions also provided a strong indication that these systems were not widely used. In the first focus group, it was described as a 'chicken and egg' problem. On one hand, researchers won't be motivated to use these metrics unless they are given recognition in assessment processes; whereas, on the other, funders and others won't consider these metrics unless they are more widely used.

OTHER ISSUES IDENTIFIED

54. In both focus group discussions, the need to adequately enable and support data use was emphasised. At present, it was noted that there was often no easy way for users to locate datasets of value. One focus group suggested that funders could potentially work towards a PubMed-like system to catalogue data resources. It was noted that if researchers were required to register data resources at an early stage, it may provide an indication of user demand – and hence lead to more cost-effective investments. It was also suggested that funders needed to do more to recognise the value of research that focuses on the secondary use and linkage of data rather than on the generation of new data, and potentially proactively stimulate such studies where appropriate. Supporting data users would in turn generate demand and incentivise data sharing, particularly if citations and other benefits flow back to the data generator.
55. Several interviewees touched on the role of ethics review groups in enabling data sharing. Overall, there was felt to be a need to enhance the capacity of ethics committees to assess the appropriateness of proposed approaches for data sharing and security. It was suggested that doing so might lead to ethics committees giving more prominence to the issue of data access and the potential benefits, which would provide an incentive to researchers to ensure they have given adequate thought to their approach for data sharing.

CONCLUSIONS AND RECOMMENDATIONS

OVERVIEW

56. It is recognised that the numbers of individuals consulted in this project was quite limited, and there is a need to apply caution in making sweeping conclusions based on these results alone. Nevertheless, some very clear and consistent messages have emerged from this work, which closely mirror the conclusions of previous policy discussions and reports on this topic.
57. The overarching view from those consulted is that, at the present time, the research culture and environment provides neither sufficient support nor adequate rewards for researchers in sharing data. Key messages were that:
- the costs of data management and sharing are often not anticipated by researchers nor resourced adequately by funders;
 - there is very little formal recognition for data sharing in key assessment processes – including in funding decisions, in academic career advancement in universities, and in the UK Research Excellence Framework;
 - the technical infrastructures and skills needed to support researchers in data management and sharing are often not in place (both at an institutional and community level);
 - the widespread lack of career opportunities for data managers in universities is a particularly critical issue in this regard, given the necessity of their skills to data-intensive science across the fields covered by EAGDA;
 - the increasing move toward large-scale collaboration in data-intensive fields creates potential barriers for early-career researchers and a need for more sophisticated approaches to assessing contributions;
 - there are very few consequences for not making data available in many cases.
58. It is important to emphasise that few, if any, of these issues are new: most have been identified and discussed at length over recent years. However, despite an overwhelming consensus that the current lack of incentives is a major barrier to enabling greater availability and use of data, very few concrete steps seem to have been initiated so far to address these issues.
59. Part of the reason for this may be that what is ultimately needed is a fundamental cultural shift in the research community to one which affords much greater value to the generation and sharing of high quality datasets. Achieving this is far from straightforward, and will require contributions from funders, institutions, journals and the broader research community.
60. It is important therefore that we identify clear and achievable actions which we could take now which will help accelerate this broader cultural change. While our recommendations are targeted primarily at UK stakeholders, we recognise that these issues are international in scope, and call on the EAGDA funders and research

community to work actively with international partners (including through the consortia of which they are part) to build incentive structures and effect cultural change.

61. The recommendations that follow are grouped into actions for funders and research leaders. We have also identified a series of issues which we believe warrant further consideration, but which either fall outside EAGDA's sphere of influence or require further exploration and scoping. While this is a complex problem, through working together, the four EAGDA funders and the researchers that they support could make a vital contribution.

RECOMMENDATIONS FOR FUNDERS

Recommendation 1: Strengthen approaches for scrutinising data management and sharing plans for their funded research – ensuring that these are resourced appropriately and implemented in a manner that maximises the long-term value of key data outputs.

62. Over recent years, the four EAGDA funders have done a huge amount of valuable work to develop policies that advance data sharing and to support the research community in increasing the accessibility of key research datasets. For example, the UK Data Service provides high quality guidance and tools for ESRC-funded researchers and other communities, together with support from a staff with extensive experience in managing complex datasets and running research and training initiatives. The MRC has likewise done a great deal of work to develop services such as the Research Data Gateway for its funded cohort and population studies, together with new guidance for researchers.

63. The evidence gathered in this report, however, provides a strong indication that there is still some way to go. In particular, while the approach of requiring researchers to submit data management and sharing plans is widely seen as the correct approach, more needs to be done to ensure that these plans are of sufficient quality and are implemented in practice.

64. Specifically, we believe that funders need to develop much more robust processes across a full cycle, which incorporates ensuring that:

- data management and sharing plans are submitted that are properly set out and costed (and which form an explicit condition of award);
- plans and the costs associated with them are evaluated adequately and resourced appropriately, taking into account the long-term costs of curation and preservation;
- the implementation of plans is tracked and information is gathered on the costs and usage of the data.

65. It is recognised that a proportionate approach is required here: data management and sharing is just one of many elements that need to be considered in assessing a grant application, and it is clear that it is a much more critical concern for some types of research compared to others. However, in those cases where a proposal is likely to generate a data resource of long-term value with potential for secondary uses, it is vital

that funders ensure the data management and sharing plans are scrutinised robustly and that there is assurance they have been implemented appropriately.

66. A critical consideration in assessing data management and sharing plans will be the availability, or otherwise, of suitable data repositories to hold, preserve and make available datasets of value. Where suitable repositories exist, there should be a clear expectation that researchers will use these unless there is a compelling reason why this is not appropriate in a particular case. Where datasets need to be maintained locally (either at a study or at an institutional level), particularly careful consideration needs to be given to how long-term curation and preservation, and their associated costs, will be provisioned. Issues around research infrastructures are discussed further below.

67. It is important to note that universities and other institutions in receipt of funding must take some responsibility for ensuring researchers plan and cost their approaches for data management and sharing appropriately. Indeed, it is in their best interests to do so. Nonetheless, funders cannot defer all responsibility to institutions and must consider what steps they can take to ensure these issues are addressed.

68. In terms of specific actions, we suggest that funders could:

- strengthen existing guidance for applicants – including through development of detailed guidance on allowable costs, and exemplars of ‘gold standard’ plans;
- provide clear guidance to reviewers and committees on the importance of this issue, proactively directing them to discuss data sharing issues where appropriate;
- engage technical experts (including data managers and data scientists) to provide specialist reviews of data management and sharing plans for proposals generating high value datasets. This could operate via an ‘adjudication’ or referral processes for proposals which have been accepted for funding, or these experts could be engaged as committee members where this is appropriate;
- build on existing web-based evaluation tools to capture information routinely on the implementation of data management and sharing plans for funded research projects – focusing in particular on research generating high value datasets.

69. A number of those consulted favoured the introduction of sanctions for non-compliance, such as withholding final grant payments (as already implemented by the ESRC). This is certainly an issue that funders should keep under review, but we believe that it is probably premature to focus on punitive sanctions at this point, particularly given that researchers may not always have been granted the resources required and may be legitimately constrained by the lack of infrastructure and support in some fields.

70. One approach funders could consider is to publish data management and sharing plans for the grants they fund (or require grant holders to do so) – this would not only place greater pressure on grant holders to ensure they are sufficiently high quality, but might also serve to aid discoverability of data for users.

Recommendation 2: Urge the UK Higher Education funding councils to adopt a clear policy at the earliest possible stage for high quality datasets that are shared with others to be explicitly recognised and assessed as valued research outputs in the post-2014 Research Excellence Framework

71. Clear and formal recognition of data outputs in the Research Excellence Framework (REF) could play a huge role in stimulating behaviour change in UK research institutions. It is suggested that the EAGDA funders should work with the UK Higher Education funding councils to advocate for this change – pressing for a clear statement at the earliest possible time. This would ideally state explicitly that high quality datasets will be considered as legitimate research outputs across all disciplines in the post-2014 REF (on an equivalent basis as research publications and other types of outputs).
72. We recognise that a key challenge here is that there are no formally agreed mechanisms to assess the ‘quality’ or ‘importance’ of a particular dataset. Therefore, in taking this issue forward, we suggest that funders commit to work with funding councils and the broader research community to build clear consensus on how quality assessments of datasets can best be implemented. As noted below, there are emerging metrics which could form part of the basis for this – but their use is not yet widespread in many research fields.

Recommendation 3: Take a proactive lead in recognising the contribution of those who generate and share high quality datasets, including as a formal criterion for assessing the track record and achievements of researchers during funding decisions.

73. In addition to the REF, there is a growing weight of opinion that research assessment procedures need to begin to consider a much wider range of metrics than is currently the case – including giving due weighting to the generation of high quality data resources. But practice has been slow to change, and publications still dominate other types of output in considering a researcher’s performance and track record. Perhaps worse still, outdated proxies for the relative quality of publications (such as journal impact factors) and for the level of contribution of an individual researcher to a paper (particularly first and last author status) are still widely used in some disciplines. The latter practice in particular is becoming increasingly questionable in an era of large-scale international collaboration and potentially damaging to the career advancement of young researchers in these fields.
74. Funders must do more to champion change. This should include providing opportunities for researchers to describe their track record in sharing data on grant applications, and providing reviewers and committee members with a clear message to take this into account where appropriate. More generally, they should be clear that high quality datasets, which are used by others, should be considered as valued research outputs. They must also ensure, when assessing publication outputs, that the contribution of researchers to a particular paper is not assessed solely on the basis of their position on author lists.

75. New metrics offer the potential to enable researchers to track the downstream use and impact of data resources, but awareness of these mechanisms in many of the communities served by EAGDA is extremely low. These tools are still in their infancy, and it may not be appropriate for funders to recommend or mandate specific approaches at this point. However, we suggest that they should actively communicate their existence, encourage researchers to assess their value, and allow researchers to cite them in grant applications.
76. Funders should also seek to publicly recognise the contribution of those who generate and share data resources. The idea of a formal prize is worth considering further. One could even envisage twin prizes – for contributions in producing data outputs, and for innovative secondary uses of datasets.

Recommendation 4: Work in partnership with research institutions and other stakeholders to establish career paths for data managers.

77. It is widely acknowledged that the skills provided by data managers in curating, managing and analysing data are becoming ever more critical to cohort and longitudinal studies, yet they often face a situation where their career prospects - at least in the university sector - are limited and their skills and contributions are not adequately recognised.
78. It is important to emphasise that the external environment is developing rapidly – with increasing attention being focused on the economic potential of the ‘big data revolution’. It is highly likely, therefore, that there will be increasing demand and new career opportunities for those with data management and analysis skills in both commercial and public sectors. It is vital therefore that the academic sector begins to recognise more seriously the value of these skills, if it is to be able to attract and retain the individuals who will be required to manage research data in the future.
79. Universities and other research institutions ultimately need to take ownership of this issue. However, given the importance of the problem, it is suggested that funders should proactively seek to drive change. It is suggested that funders should consider:
- providing targeted funding for data managers, perhaps through a dedicated training fellowship scheme;
 - exploring the potential to formalise the role of ‘data manager’ or ‘data scientist’ through professional accreditation (working with the UK Digital Curation Centre and other key partners);
 - setting good practice standards and expectations – for example, requiring applicants and institutions to have a clear commitment to career development of their research teams;
 - working to raise the status of data managers - engaging them as technical experts in funding processes (as suggested above), or providing them with a formal role in training scientists in data management skills.

Recommendation 5: Ensure key data repositories serving the data community have adequate funding to meet the long-term costs of data preservation, and develop user-friendly services that reduce the burden on researchers as far as possible.

80. For some areas covered by EAGDA, well-established data repositories exist – such as the UK Data Service for social science data, and the European Genome and Phenome Archive (EGA) and other European Bioinformatics Institute (EBI) resources for genomic and associated phenotypic studies. For many epidemiological studies, however, centralised repository infrastructure is not yet available and data are typically held by the research team or institution.
81. Data repositories play a critical role for the communities they serve: not just in ensuring datasets of value are preserved in a useable and accessible form, but also in developing and implementing standards, and in reducing the burden on individual researchers of managing and sharing data. Typically, however, these key resources lack sustained long-term funding – which threatens the key services they provide. The challenge of sustaining data infrastructure is a significant one, which clearly goes beyond questions of incentives and is being addressed in other UK and international forums (including for example in the ELIXIR project at EU level). Nonetheless, these enabling infrastructures are such an essential prerequisite for building a culture that supports wider data access.
82. One issue that was highlighted consistently by those consulted in this study was that many of the existing repositories were not as user-friendly as they could be. This concern was highlighted particularly strongly in relation to the EGA, which it was perceived is not sufficiently well-funded and is difficult to use – creating a barrier for researchers who wished to make data available. We suggest that funders should actively explore how they could work with the EGA and other key resources to help build services that reduce the burden on data generators as far as possible.
83. It is also particularly critical that we do not lose sight of those datasets that are not served by existing repositories. Over the longer term, funders should actively consider whether existing repository infrastructures could be expanded, or new repositories developed, to cover these data. We would emphasise that, while there are considerable benefits of using centralised repositories wherever that is possible and appropriate, some classes of data are better managed via disaggregated repositories that can be updated and curated by the original data generator and can be shared and co-analysed in a federated manner.
84. In the more immediate term, as noted above, funders must ensure that the costs of preservation and access have been anticipated and resourced. Arrangements where costs are borne at least in part by data users may be appropriate in some cases. A pressing related challenge is to ensure that these datasets are readily discoverable by potential users. Initiatives such as MRC's Research Data Gateway are a valuable step towards this.

RECOMMENDATIONS FOR RESEARCH LEADERS

Recommendation 6: Adopt robust approaches for planning and costing data management and sharing plans when submitting funding applications.

85. While funders must ensure data management and sharing plans are reviewed adequately and required costs are met, research leaders must also accept responsibility for ensuring that they plan their approaches for data management and sharing robustly, adopt recognised best practice standards, and carefully consider the resources and skills they will need to deliver on their plans appropriately.

Recommendation 7: Ensure that the contributions of both early-career researchers and data managers are recognised and valued appropriately, and that the career development of individuals in both roles is nurtured.

86. Research leaders have a core responsibility to all members of their teams to ensure that they are able to develop in their roles, and that they receive due recognition for their contributions. Following on from the point above, it is vital that study leaders are realistic and transparent about the skills they require to deliver a particular project. Specifically, the burden of data management must not be unduly passed onto postdoctoral researchers (unless they have an explicit desire to develop in this direction and to take on a data scientist type role). If study leaders require specialist data management expertise, then this is what they should request.

87. The increasing prevalence of large multi-author papers in large-scale cohort and longitudinal studies can be seen as a hindrance by early-career researchers, although the situation is not clear cut – in that many also feel they have benefited from taking part in large-scale consortia. Overall, it is not clear from this research whether it is severe enough to discourage researchers from pursuing research careers in these fields.

88. There are potential lessons that could be learned from other disciplines here - particularly particle physics and astronomy, which has developed a system of researchers providing a qualitative description of their contributions to key outputs in assessment processes. Although still in their infancy, there are also emerging systems to assign contributor roles to authors in papers which could also have a key role. It is suggested that studies should actively seek to adopt these where they can add value.

89. These systems could also help to provide recognition for data managers, as could the emergence of new data journals – where, for example, a data manager could potentially be included as a lead author on a data paper. However, it is important to note that standard publication outputs may not be a key motivator for some data managers, nor should data managers be assessed against these traditional academic outputs. More generally, we suggest that study principal investigators should actively embrace ways in which they can provide due recognition and credit to data managers in their teams – including through advocacy at institutional level to enhance their career opportunities.

Recommendation 8: Develop and adopt approaches that accelerate timely and appropriate access to key research datasets.

90. In seeking to maximise the value of their data, studies should actively consider ways in which data outputs can be made available to legitimate research users at an early stage. At the current time, however, it is recognised that researchers may be unwilling to do more in this regard without some steps to protect their ability to secure publications based on the data.
91. The publication moratorium system is thought by many to have worked reasonably effectively in the genomics community in enabling the early sharing of datasets, and is potentially an approach that could be considered in other fields. However, there may well be alternative mechanisms that could be considered, including formal stipulations in data access agreements. Ultimately, the hope would be that the research culture would shift to a point where the benefits flowing back to data generators would support early sharing of data, without the need for moratoria and other restrictions. However, there is probably some way to go before we reach this point.

Recommendation 9: Champion greater recognition of data outputs in the assessment processes to which they contribute.

92. Study leaders and other researchers involved in data intensive science could play a major role in stimulating cultural change through acting as champions in their institutions and in the research assessment processes to which they contribute. Many of those who were consulted in this study and who highlighted the need for greater incentives are actively serving as members of funding committees, institutional promotion boards and REF panels. They should, wherever possible, ensure that they adopt these principles in practice and work to shift attitudes among their peers.

ISSUES FOR RESEARCH INSTITUTIONS AND JOURNALS

93. There are several other critical elements required to stimulate cultural change and build the incentive structures required to support data sharing – in particular, **both research institutions and research journals have a critical role to play**. While it is not the role of EAGDA to make recommendations to these communities, we would urge these stakeholders to work with funders and the wider research community in helping to address the issues raised in this report.
94. It is absolutely clear that funders and research leaders will not be successful in advancing wider cultural change, without support from universities and other research institutions. We would call on these institutions to work proactively to:
- set clear policies on data management and preservation (as many already have), and communicate these to their staff;

- support their researchers in planning and costing their approaches for managing research data;
- act urgently to develop defined career structures for data managers and ensure they are recognised as key members of research teams;
- Implement institution-wide approaches to train researchers in core data management skills– utilising the skills of data managers wherever possible;
- ensure data outputs are given greater weighting in career advancement processes;
- consider the broader need to enhance operational and project management skills in research teams.

95. Scientific journals could make a major impact in driving culture change through having robust data sharing policies for the sharing of data underlying published papers. In December 2013, PLOS announced a new policy requiring data availability²¹, and it is to be hoped that other journals will follow their lead. Journals also have a key role in promoting mechanisms to allow the impact and citation of datasets to be tracked, and for author contributions to be determined. We call on journals to:

- set in place consistent mechanisms to better attribute the contribution of authors to publications in data-rich fields;
- develop clear requirements for the citation of datasets used in published research;
- ensure that datasets reported in published papers have associated details on how they can be accessed, and require their deposition in externally-accessible repositories where these exist.

REFERENCES

1. Science as an open enterprise (The Royal Society, June 2012) [link](#)
2. Walport, M.J. and Brest, P. (2011). Sharing research data to improve public health. *Lancet* 377: 537-539
3. Large-scale data sharing in the life sciences: data standards, incentives, barriers and funding models (Digital Archiving Consultancy, 2005). [link](#)
4. Dealing with data: roles, rights, responsibilities and relationships (UKOLN 2007) [link](#)
5. Shared responsibilities in sharing research data: policies and partnerships: report of an ESF-DFG workshop (European Science Foundation 2007) [link](#)
6. To share or not to share: Publication and quality assurance of research data outputs (RIN, 2008) [link](#)
7. Piwowar, H.A. and Vision, T.J. (2013): Data reuse and the open data citation advantage. *PeerJ* (DOI 10.7717/peerj.175) [link](#)
8. Tenopir, C. et al (2011): Data sharing by scientists, practices and perceptions. *PLOS One* 6 (6) [link](#)
9. Hartter, J. et al (2013): Spatially explicit data: stewardship and ethical challenges in science. *PLOS Biology* 11 (9) [link](#)
10. European landscape study of research data management (SURF 2013) [link](#)
11. Keeping research data safe 2 (HEFCE, 2010) [link](#)
12. Toronto International Data Release Workshop Authors (2009). Prepublication data sharing. *Nature* 461, 168-170
13. Joly, Y. et al (2012): Open science and community norms: data retention and publication moratoria policies in genomics projects. *Medical Law International* 12, 92-120
14. Enabling the citation of datasets generated through public health research (Wellcome Trust, 2012) [link](#)
15. Costas, R., Meijer, I., Zahedi, Z. and Wouters, P. (2013). The Value of Research Data - Metrics for datasets from a cultural and technical point of view (a Knowledge Exchange Report) [link](#)
16. Making Data Count: Research data availability and research assessment – workshop report (Knowledge Exchange, April 2013) [link](#)
17. Draft report on incentives and rewards in the field of biomedical research databases (GEN2PHEN, August 2010) [link](#)
18. Helping to open up: improving knowledge, capability and confidence in making research data more open (Research Information and Digital Literacies Coalition, 2013) [link](#)
19. A comparative Study of International Approaches to Enabling the Sharing of Research Data (JISC, 2008) [link](#)
20. ENCODE data use policy for external users (March 2014) [link](#)
21. Data Access for the Open Access Literature: PLOS's Data Policy (December 2013) [link](#)

Cancer Research UK
The Angel Building
407 St John Street
London EC1V 4AD
T +44 (0)20 3469 8360
E publicaffairs@cancer.org.uk
www.cancerresearchuk.org

Economic and Social Research Council
Polaris House
North Star Avenue
Swindon SN2 1UJ
T 01793 413000
E comms@esrc.ac.uk
www.esrc.ac.uk

Medical Research Council
Polaris House
North Star Avenue
Swindon SN2 1FL
T 01793 416200
E corporate@headoffice.mrc.ac.uk
www.mrc.ac.uk

Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE, UK
T +44 (0)20 7611 8888
F +44 (0)20 7611 8545
E contact@wellcome.ac.uk
wellcome.ac.uk

This work is © the Wellcome Trust and
is licensed under Creative Commons
Attribution 2.0 UK.