

Data Management Plans: Recommendations

Introduction

1. There is considerable interest from research funders in how to maximise the value of research outputs, including data. The UK Concordat on Open Research Data¹ is the latest UK initiative to push for research data to be more discoverable, accessible and useable by others beyond the study team and immediate collaborators.
2. One mechanism for operationalising this policy priority is through the requirement that grant applications include a data sharing and management plan. Funders vary in the extent of their requirements but all emphasise that data should be made available for reuse as far as possible.
3. Funders are keen to support data access for several reasons. Increasing data availability will:
 - enable research data to be verified and built upon;
 - facilitate novel research questions previously unanticipated lines of research;
 - reduce unnecessary duplication;
 - reduce the need for costly primary data collection (and therefore make investments more cost-effective);
 - speed the advancement of scientific knowledge.
4. There can also be significant and tangible impacts on health, for example, rapid data sharing in public health emergencies has the potential to improve clinical and public health decision-making in real-time in an epidemic situation.
5. Enabling data access also has benefits in helping shift entrenched cultures that are potentially damaging: requirements to make data available can act as a lever to help break territorialism over data. There is a great deal of exciting potential to be unlocked in some datasets, both for its intrinsic scientific value and for enabling junior researchers to develop, explore innovative ideas and progress their careers. The arguments in favour of enabling access to research data have been well-established for several years², but academic research in practice has, in some fields, been slow to adapt.
6. In the US, there is a significant focus on ‘team science’ and collaboration in many well-established consortia. It is becoming more normalised for data to be shared extensively. The institutional, funding and legal infrastructure necessary to facilitate data sharing is comparable across the UK, US and EU³, and the richness of potential data resources generated by UK scientists (especially in epidemiology and social science) means there is significant global demand for UK data.
7. The flipside of this positive drive towards enabling data to be used more widely is that failure to do so may result in data requestors seeking other routes to ensure access to data. In the UK, universities do fall under Freedom of Information Act (2000)/ Protection of

¹ <http://www.rcuk.ac.uk/documents/documents/concordatonopenresearchdata-pdf/> [Accessed 28 July 2016]

² See for example ‘Science as an Open Enterprise’ <https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/> Royal Society (2012)

³ One particular challenge in this area is the appropriateness of the US ‘open science’ data sharing agenda for human health data, given the need for protections and controls on its use.

Freedoms Act (2012)⁴ requirements and researchers may be forced to hand over data for release into the public domain against their wishes, as was the case in the recent PACE tribunal⁵.

8. In this context, failing to plan adequately for, and manage, data outputs can generate reputational risks for both funders and researchers. If UK researchers are unequipped to rapidly share their data they could be at risk of being perceived as not working collectively together to maximise the research potential of their resources, or may be forced to release data under unfavourable conditions, for example through an FoI request. Expectations need to be managed across borders, with clarity about what data may be made available, when and under what conditions.

9. This requires a shift in scientific culture, which is a complex, long-term process. EAGDA funders have a real opportunity to act jointly to address these challenges, breaking down disciplinary siloes across areas of their funding to champion the value and benefits of access to data.⁶ Funders acting independently of one another will retrench existing siloes in policy, language and practice: only through a co-ordinated, strategic effort to work together across funders will approaches to managing data in practice shift over time.

Policies to Practice

10. Including data sharing and management plans ('data plans') as part of a funding application is one tool funders can use to drive a specific change within their research communities. Data plans afford applicants the opportunity to think carefully and creatively about how they should manage the outputs of their research.

11. To exercise this function effectively, data plans must form an integral part of the funding application process, and there must be mechanisms in place to assess such plans, provide adequate resources to carry out such the relevant activities, and to monitor progress as they are carried out.

12. The implementation of funder policies is crucial to realising this potential. However, the extent to which this policy requirement is being taken up in practice is unclear. If data plans are not adequately assessed in the first instance and monitored, with funders being clear that they care about adherence to them, there is a risk that data management remains an afterthought.

13. The evidence gathered by EAGDA and documented in the annexes to this paper (**Annexes 1-6**) suggests that, despite a similarity in stated policies, there is significant variation within and between funders in when and how data plans are taken into account in the funding decision process and follow up – and very little transparency for applicants about what the funders do in practice in these respects.

⁴ The Protection of Freedoms Act (2012) amends the FoI Act to require public authorities to release datasets in an electronically useable format.

⁵ See for coverage: <http://www.bmj.com/content/354/bmj.i4614>

⁶ The Protection of Freedoms Act (2012) established the mandate to enable secondary use of data generated by public bodies, signalling a shift towards openness that academic research should strongly advocate.

Recommendations

EAGDA makes the following specific recommendations to funders, divided into two types:

1. Recommendations for specific actions that can be implemented across scientific domains independently of the existing structures and processes funders have.
2. Indications of issues pertaining to existing structures and processes that we recommend funders consider, either jointly or separately, if they wish to effectively promote and enhance data sharing in their research communities.

1. Actions for funders

1.1. *Make the case clearly and set expectations*

Funders should work in a joined up, coherent way to provide detail beyond high-level policy statements about why they consider data sharing to be important. This statement or argument should be available to funding committees, peer reviewers, funder staff and the research communities they support. It should include:

- What *specific* benefits they see resulting from increasing the availability, accessibility and usability of research data;
 - this should include reciprocity for data producers, setting the expectation that data users return enriched data to the resource;
- An indication of what credit or recognition they will give to researchers who make data available, accessible and useable;
- Expectations for data sharing with international partners and consequences for not adhering to good practice;
- The opportunity costs of failing to deliver; and
- In what fields, and at what scale or type of study, they consider data sharing to be a priority.

1.2. *Demonstrate transparency and accountability*

Funders should:

- Include the data plan, or a short summary of the commitments contained within it, alongside the other publicly available summary details of grants awarded.
- Establish clear processes for how data plans will be assessed, at what stage in the application process, by whom, and what value and weight will be placed on them in funding decisions. This information should be provided to prospective applicants, in-house staff handling grant applications, funding committees and peer reviewers.
- Develop internal guidance for staff so that they can consistently identify which applications require support to develop a data plan at an early stage and signpost applicants to relevant resources where needed.
- Clearly state the consequences of failing to adequately plan for or adhere to plans for data sharing, and what they will do to evaluate this.
- Mandate data plans to be published, as documents that evolve over the course of a grant. Full details of plans should be made publicly available along with research

protocols, to enable data users to find out what researchers have committed to doing and over what timescale. In the longer-term, plans should be machine readable.

- Require self-reporting on data management activity via ResearchFish, annual review reporting or at the grant renewal stage. These reports should be taken into account by funders during the review process.
- Put mechanisms in place to enable funders to report in two years' time on whether and how research practice has changed as a result of implementing these recommendations.

1.3. Develop guidance for funding committees

Funders should develop guidance for their respective funding committees stating what (if anything) they expect from them in terms of assessing data plans. Such guidance will only be effective if accompanied by a clear commitment from funders on recommendations 1.1 and 1.2 above.

- The guidance should be clear on what committees are expected to consider and where in the funding process judgements about the quality of data plans will be made.
 - It should be explicitly stated whether or not data plans must meet a threshold of quality and what impact (if any) failure to meet this threshold should have on the funding decision.
 - Specific guidance will necessarily vary across committees and scientific disciplines. It will need to be proportionate and adaptable as cultures of data sharing shift in different fields over time.
- Funders should provide, as needed, up to date information for committees on the kinds of tools and infrastructure that are available for data sharing in different fields, to encourage committees to challenge applicants and be aware of best practice.⁷

2. Issues for funders to consider

Data sharing is a complex and heterogeneous issue that is evolving constantly, with standards and best practice developing in some areas but a continuing lack of consensus in others. Specific funders also have different policies and processes, different scientific domains to cover and will not all benefit from the same approach to implementing their policies.

At this point, on the basis of the research that has been undertaken for this report, the view of EAGDA is that is neither appropriate nor feasible to make further specific or detailed recommendations that will be applicable to all of its funders. EAGDA nonetheless urges them to consider strategically the following issues.

2.1. Review policy implementation

Funders data sharing policies all champion the value of data availability and access, but there appears to be little joined-up funder evidence-gathering and analysis of whether and

⁷ For example, through promoting adoption of the FAIR (Findable, Accessible, Interoperable, Reusable) principles: <http://www.nature.com/articles/sdata201618>

how these policies are changing researcher behaviours and culture in practice. This evidence gap needs to be acknowledged, with funders taking responsibility for addressing it.

Without evidence about what is happening to data plans in practice it will not be possible to develop measures to identify obstacles, assess progress or promote good practice. Funders should jointly consider how they could track and assess planning for data management, including:

- Which applications include data plans and their level of quality;
- Whether there is a clear, consistent internal view about which applications should include data plans (Wellcome), or which applications will generate data that may be particularly valuable for secondary use and therefore may warrant particular scrutiny (CRUK, ESRC and MRC);
- Where in the grant process any issues with data plans are picked up and addressed, and how.

2.2. Consider data as part of broader outputs planning

Data is one form of research output that funders are encouraging researchers to make more openly available. Other valuable outputs include software, IP and publications. It may be valuable for applicants to consider how all of their outputs are going to be managed and shared: data access and management should not necessarily be planned for separately from research outputs more broadly.

Funders should consider developing policies or guidance that encourage applicants to take a comprehensive approach to planning for what they will do with all of the outputs of their research, including data. This approach may help address the concern that funders are putting too much emphasis on data sharing at the expense of other aspects of the research enterprise. It will also encourage researchers, where relevant, to think about such issues as how to balance IP implications with post-competitive data sharing at the outset of their research planning.

Funders should consider whether downstream uses of data that are shared can be tracked, in order to ensure that the scientific value and impact of the original investment can be accurately measured and that the original project team receive appropriate credit. Such mechanisms would ideally need to cover all subsequent research and analysis not just that funded by EAGDA funders.

2.3. Establish an accurate guide to the costs of data management

Funders should attempt to understand and establish benchmark costs for data management and access for different types of data and with different levels of governance and control. To do this it will be necessary to gather more data on grantees' spend on data management and access. There are existing tools for analysis that could be used to provide approximations about the time, resources and infrastructure costs that are needed to support data management. Clarity on this for applicants and funding committees would enable proposed costs in data plans to be reasonably assessed.

Funders should ensure they are able to accurately identify the costs associated with data access, governance and management in grants that they award. This could enable more judicious use of funds to supporting coherent, harmonised data management practices, for example, across several grants within an institution.

Funders should also engage with institutions to clarify where they consider responsibility for long-term data management and storage to lie: with institutions, PIs, and/or data repositories. The costs associated with individual grants should reflect these arrangements, with additional resource being allowed for in the absence of established institutional or other appropriate infrastructure to support data management.

2.4. Determine how data plans should be monitored

Funders should work together to determine how the implementation of data plans should be monitored and adherence checked, and what types of research this would be appropriate for.

Several mechanisms in addition to annual reporting might be explored, including:

- **Independent ombudsman:** a data ombudsman arbitrates when there are conflicts about the availability of data, working across scientific disciplines and advocating for best practice.
- **Funder auditing:** in-house regular 'dipstick' checks of adherence to data plans.
- **Track records:** data sharing track records included in future funding applications. These should pick up researchers' efforts to make data available and accessible where appropriate, and should not focus on publication outputs resulting from secondary data use.

Funders should also consider whether and how to develop metrics for:

- Monitoring adherence to data plans once agreed and adequately resourced;
- Checking that internal funder processes are picking up the right applications to support with their data management plans during the application or renewal.

2.5 Benchmark now to evaluate change over time

Funders should put mechanisms in place now to benchmark current data management and access practices across their funded research communities. This will enable evaluation in years to come and to identify what changes have had an impact in practice. Such evaluation will need to be done on an ongoing basis if funders want to maximise the value of data generated through funded research.