Wellcome Trust: Wellcome Trust template

Data outputs

i. What data outputs will your research generate and what data will have value to other researchers?

Guidance:

Researchers should maximise access to research datasets of value to the wider research community in a timely and responsible manner. Any data that is shared should be of a sufficiently high quality that it will have value to other researchers and should be provided in a format that enables it to be used effectively.

We recognise that in some cases it may not be appropriate for researchers to share their data. However, if your research meets the criteria for requiring a data management and sharing plan but you are intending not to share your data, the reasons for this must be clearly justified.

Data should be shared in accordance with recognised data standards where these exist, and in a way that maximises opportunities for data linkage and interoperability. Sufficient metadata must be provided to enable the dataset to be used by others. Agreed best practice standards for metadata provision should be adopted where these are in place.

When developing data management and sharing plans, researchers should therefore consider and briefly describe:

- what types of data the proposed research will generate?
- which data will have value to other research users and could be shared?
- what data formats and quality standards will be applied to enable the data to be shared effectively?

Data sharing

ii. When will you share the data?

Guidance:

All data management and sharing plans must state clearly the timescales over which datasets of value will be shared. Such timescales should take account of any recognised standards of good practice in the applicant's research field.

In considering the timescales that are appropriate, the Trust recognises fully that data generators have the right to a reasonable (but not unlimited) period of exclusive use for the research data that they produce.

As set out in our guidelines on good research practice, all grant holders must ensure as an absolute minimum that the data underpinning research papers are made available to other researchers on publication, providing this is consistent with any ethics approvals and consents which cover the data and any intellectual property rights in them.

In cases where the creation of a database resource is the primary goal of a Trust-funded activity, we would normally expect the data to be made widely available to user communities at the earliest feasible opportunity.

In line with the Fort Lauderdale Principles and subsequent Toronto statement on pre-publication data sharing, the Trust also encourages timely and responsible pre-publication data sharing for research that might constitute a "community resource" (i.e. those that have the characteristics set out in point 7 above).

Where appropriate, researchers may use publication moratoria to facilitate pre-publication sharing of data with other researchers, while protecting their right to first publication. Any such restrictions on data use should be reasonable, transparent and in line with established best practice.

Illustrative examples of timescales for data sharing are provided to help demonstrate different models that have been adopted and may be considered as examples of good practice in the field of large-scale genetics and genomics studies.

iii. Where will you make the data available?

Guidance:

Researchers should deposit data in recognised data repositories where these exist for particular data types, unless there is a compelling reason not to do so. Further information on repositories that may be appropriate.

If the intention is to create a tailored database resource or to store data locally, researchers should ensure that they have the resources and systems in place so that the data are curated, secured and shared in an

way that maximises its value and safeguards any associated risks.This includes consideration of how data held in this way can be effectively linked and integrated with other datasets to enhance its value to users.

iv. How will other researchers be able to access the data?

Guidance:

Data should be made available to other researchers with as few restrictions as possible. Where a managed access process is required - for example, where a study involves potentially identifiable data about research participants - the access mechanisms established should be proportionate to the risks associated with the data, and must not unduly restrict or delay access. Any managed access procedures that are proposed must be described clearly as part of your data management and sharing plan.

Depending on the study, it may be appropriate to establish a graded access procedure in which less sensitive data (e.g. anonymised and aggregate data) are made readily available, whereas applications to access to more sensitive datasets are subject to a more stringent assessment process.

Any managed access procedures should be consistent and transparent. In cases where a Data Access Committee is required to assess applications to access data, the composition of such Committees should include individuals with appropriate expertise who are independent of the project.

Where appropriate, the Trust would encourage those generating datasets that are likely to be of significant value to other researchers to publish a 'marker paper' or other form of publication, which enables data users to formally cite their usage of the resource.

Where a database resource is being developed as part of a funded activity, researchers should take reasonable steps to ensure that potential users are made aware of its availability. These should be outlined briefly in your plan.

v. Are any limits to data sharing required? For example to safeguard research participants or to gain appropriate intellectual property protection.

Guidance:

For some research, delays or limits on data sharing may be necessary and appropriate to safeguard research participants or to ensure intellectual property protection is gained. Any such restrictions should, however, be minimised as far as feasible and set out clearly in data management and sharing plans where these are required.

Safeguarding Research Participants

For research involving samples or information pertaining to human subjects, data must be managed and shared in a way which is fully consistent with the terms of the consent under which samples and data were provided by the research participants.

For prospective studies, consent procedures should include provision for data sharing in a way that maximises the value of the data for wider research use, while providing adequate safeguards for participants. As part of the consent process, proposed procedures for data sharing should be set out clearly and current and potential future risks associated with this explained to research participants.

In designing studies, researchers must ensure that they have appropriate systems to protect the confidentiality and security of data pertaining to human subjects, and minimise any risks of identification by data users. This can be achieved through the use of appropriate anonymisation procedures and managed access processes. Such systems should be sufficient to safeguard participants, but proportionate to the level of sensitivity of the data and associated risk. They should not unduly inhibit responsible data sharing for legitimate research uses.

Intellectual Property

In line with our policy on intellectual property and patenting, we expect our funded researchers to ensure that any intellectual property in the outputs of their research is suitably protected and managed in a way that best enables the use of that knowledge for ultimate health benefit.

Delays or restrictions on data sharing may be appropriate to gain intellectual property protection or to further development of a technology for public benefit. As noted above, any such limits should be minimised as far as is feasible.

Data preservation

vi. How will you ensure that key datasets are preserved to ensure their long-term value?

Guidance:

Researchers must consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. If the proposal is to create a bespoke data resource or to store data locally rather than to use a recognised data repository, data management plans should state clearly how the applicant expects that the dataset will be preserved and shared when the period of grant funding comes to an end.

The Trust is happy to discuss issues relating to longer-term preservation and sustainability with researchers so as to help provide the support required to maximise the long-term value of key research datasets.

Required resources

vii. What resources will you require to deliver your plan?

Guidance:

In preparing data management and sharing plans, researchers should consider carefully any resources they may need to deliver their plan. Where dedicated resources are required, these should be outlined and justified as part of the plan.

Issues to consider include:

- People and skills: is there sufficient expertise and resource in the research team to manage, preserve and share the data effectively? Is additional specialist expertise (or training for existing staff) required? If so, how will this be sourced?
- Infrastructure: are there appropriate computational facilities to manage, store and analyse the data generated by the research?
- Tools: will additional computational facilities and resources need to be accessed, and what will be the costs associated with this?