Wellcome Trust: Wellcome Trust Template

Data and software outputs

The data and software outputs your research will generate and/or re-use

Guidance:

Consider and briefly describe:

- the types of data and software the proposed research will generate
- which data and software will have value to other research users and could be shared.

We recognise that in some cases it may not be appropriate for researchers to share data and software outputs (for example, for ethical or commercial reasons). If you don't intend to share outputs, you must justify your reasons.

Software should be shared in a way that allows it to be used effectively, and we encourage you to provide appropriate and proportionate documentation for the user community.

We encourage you to share null and negative findings and data, as well as data supporting new findings, where this may have value to the community. This helps to avoid unnecessary waste and duplication.

When existing data/resources are being re-used as part of the funding activity, you should consider:

- how existing data/resources will be accessed
- if there are any constraints on re-use of existing data
- how data provenance will be documented.

The metadata and documentation that will accompany the outputs

Guidance:

Data should be shared in line with recognised data standards, where these exist, and in a way that maximises opportunities for data linkage and interoperability. FAIRSharing(opens in a new tab) is one directory of available data standards.

You should:

- provide sufficient, high quality metadata to allow the dataset to be discovered, interpreted and used by others
- adopt agreed best practice standards for metadata provision, where these are in place.

When you intend to share your data and software

Guidance:

You must specify the timescale for sharing datasets and software, using any recognised standards of good practice in your research field.

Researchers have the right to a reasonable (but not unlimited) period of exclusive use of the research data and software they produce.

As a minimum, you should make the data and software underpinning research articles available to other researchers **at the time of publication**, providing this is consistent with:

- any ethics approvals and consents that cover the data
- reasonable limitations required for the appropriate management and exploitation of IP.

Please read our requirements for publishing Wellcome-funded research papers [PDF 1MB] for more information.

Where research data relates to a public health emergency, quality-controlled data must be shared as rapidly and openly as possible. This is in line with the joint statement on data sharing in public health emergencies and GLOPID-R principles for data sharing in public health emergencies.

We encourage researchers to consider opportunities for timely and responsible pre-publication sharing of datasets and software. Where appropriate, you may use publication moratoria to enable pre-publication sharing with other researchers, while protecting your right to first publication.

Any restrictions on data and software use should be reasonable, transparent and in line with established best practice in the respective field.

Where your data and software will be made available

Guidance:

You should deposit data in recognised data repositories for particular data types where they exist, unless there's a compelling reason not to do so. The FAIRSharing(opens in a new tab) and Re3Data(opens in a new tab) sites provide lists of data resources, and Wellcome Open Research maintains a curated list of approved repositories(opens in a new tab) suitable for Wellcome-funded research.

Where there is no recognised subject area repository available, we encourage researchers to use general community repositories and resources, such as Dryad(opens in a new tab), FigShare(opens in a new tab), the Open Science Framework(opens in a new tab) or Zenodo(opens in a new tab).

If you intend to create a tailored database resource or to store data locally, you should ensure that you have the resources and systems in place to curate, secure and share the data in a way that maximises its value and guards against any associated risks.

You need to consider how data held in this way can be effectively linked to and integrated with other datasets to enhance its value to users.

For software outputs, use a hosting solution that exposes them to the widest possible number of users. GitHub(opens in a new tab) allows revision control and collaborative hosting of project code for software development, with associated archiving of each release in Zenodo. A suitable revision control system and issue tracker should be in place before programming work begins. This should be available for all members of the research team.

How your data and software will be accessible to others

Guidance:

Your plan should set out clearly:

- how potential users will be able to discover, access and re-use data or software outputs
- any associated terms or conditions.

Enabling discovery

Where a data or software resource is being developed as part of a funded activity, you should take reasonable steps to ensure that potential users are:

- made aware of its availability
- updated on significant revisions and releases.

Your plan should outline your approach for maximising the discoverability of your data or software.

We encourage all researchers to use digital object identifiers (DOIs) or other persistent identifiers for their data and software outputs, to enable their re-use to be cited and tracked.

The DataCite initiative(opens in a new tab) provides a key route through which DOIs are assigned to datasets. Many repositories assign DOIs on deposition.

Where appropriate, you may also publish an article describing dataset or software output to help users discover, access and reference the resource. You can use venues such as Scientific Data(opens in a new tab), Giga Science(opens in a new tab) and Wellcome Open Research(opens in a new tab).

Access procedures for data

Where a managed data access process is required – for example, where a study involves identifiable data about research participants – the access mechanisms should be proportionate to the risks associated with the data. They must not unduly restrict or delay access.

You must describe any managed access procedures in your outputs management plan. It should be consistent and transparent and documented clearly on your study website.

Depending on the study, you may want to establish a graded access procedure where less sensitive data – for example, anonymised and aggregate data – is made readily available, and more sensitive datasets have a more stringent assessment.

Where a Data Access Committee is needed to assess data access requests, the committee should include individuals with appropriate expertise who are independent of the project.

The Expert Advisory Group on Data Access has set out key principles for developing data access and governance mechanisms, to which applicants should refer.

Open software and database licences

If you're sharing your output through a repository, the terms by which you do so are likely to be set by the repository itself. If you're sharing directly with the research community, you need to consider the most appropriate way to do so, for example by an appropriate open licence or public domain dedication.

For data, we recommend Creative Commons licences such as CC0 or CC BY. For software, the Open Source Initiative(opens in a new tab) provides access to a range of open software licences, such as the GNU General Public Licence, Apache Licence, and the MIT Licence. Where possible, you should select one of these standard licences (rather than using a bespoke licence).

You must make sure it's clear which licence has been applied, so that users can see whether the data or software is accessible and on what terms.

Whether limits to data and software sharing are required

Guidance:

For some research, delays or limits on data sharing may be necessary to safeguard research participants or to ensure you can gain IP protection.

Restrictions should be minimised as much as possible and set out clearly in your outputs management plans, if required.

Safeguarding research participants

For research involving human subjects, data must be managed and shared in a way that's 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. For more information about consent for data sharing, go to the UK Data Service(opens in a new tab). Procedures for data sharing should be set out clearly, and current and potential future risks explained to participants.

When designing studies, you must make sure that you protect the confidentiality and security of human subjects, including through appropriate anonymisation procedures and managed access processes.

Clinical trials

For clinical trials you should mention specifically how you will share individual-level patient data. This should include:

- a plan to seek patient informed consent that allows data to be shared in the way outlined
- the level of identification risk and method of de-identification you will adopt
- the repository you plan to use
- any managed access arrangements, such as a data access committee.

Intellectual property (IP)

Delays or restrictions on data or software sharing may be appropriate to protect and use IP in line with our policy on intellectual property and patenting. If this applies, you should only share data or software when it no longer jeopardises your IP position or commercialisation plans.

Your proposed approach for identifying, protecting and using IP should be set out as described in the IP section of this guidance below.

How datasets and software will be preserved

Guidance:

You need to consider how datasets and software that have long-term value will be preserved and curated beyond the lifetime of your grant.

If your proposal is to create a bespoke data or software resource, or to store data or software locally rather than to use a recognised repository, your plan should state how you expect to preserve and share the dataset or software when your funding ends.

Research materials

What materials your research will produce and how these will be made available

Guidance:

Your plan should identify any significant materials you expect to develop using Wellcome funding,

which could be of potential value as a resource to other researchers.

You should identify in your plan how the materials will be made available to potential users. For example, by:

- depositing in a recognised collection such as ECACC(opens in a new tab)
- licensing to a reputable life science business partner who can handle advertising, manufacture, storage and distribution.

If the material is highly specialised and the potential number of users is so small that commercial partners cannot be found, distributing samples yourself to other researchers who have asked for them, may be an acceptable plan. However, where possible, you should find a more sustainable long-term solution that doesn't put an undue burden on you or your institution.

When dealing with commercial entities, you should retain the right to produce the research materials yourself, and to license others to do so, if your chosen commercial partner is unable or unwilling to continue supplying them to the research community.

Whilst your institution may generate reasonable revenue from commercialising research materials, the primary driver should not be revenue generation. You should ensure that your research materials are made available to the wider research community and thereby advance the development of health benefits.

Resources required

You should consider what resources you may need to deliver your plan and outline where dedicated resources are required.

Guidance:

Examples of resources you can request include:

People and skills

- support for one or more dedicated data manager or data scientist (full- or part-time)
- data and software management training for research or support staff that are needed to deliver the proposed research.

We don't usually consider costs for occasional or routine support from institutional data managers or other support staff.

Storage and computation

- any dedicated hardware or software that is required to deliver your proposed research
- the cost of accessing a supercomputer or other shared facilities.

We would usually expect costs associated with routine data storage to be met by the institution. We will only consider storage costs associated with large or complex datasets which exceed standard institutional allowances.

Access

- the reasonable costs of operating an access committee or other data access mechanism over the lifetime of the award
- the costs of preparing and sharing data, software or materials with users (and whether cost-recovery mechanisms will be used)
- the costs of ingesting secondary data, code or materials from users
- costs associated with accessing data, software or materials from others researchers that you need to take forward your proposed research.

Deposition and preservation of data, software and materials

- ingestion or deposition costs to recognised subject repositories for data, code and materials
- the costs for data or code deposition in unstructured repositories (eg FigShare, Dryad and Zenodo) where no recognised subject repository exists.

If no repository is suitable, we may consider ingestion costs for institutional repositories.

We don't usually consider estimated costs for curation and maintenance of data, code and

materials that extend beyond the lifetime of the award. But we’re willing to discuss how we can help support the long-term preservation of very high-value outputs on a case-by-case basis.

Intellectual property

What IP your research will generate

Guidance:

Your plan should describe any significant IP that is likely to arise during your research. You should identify what processes you have in place to identify and capture this IP, as well as any unanticipated discoveries or inventions that result from your work.

How IP will be protected

Guidance:

You should describe if and how you will protect significant Wellcome-funded IP. For example, if you’re registering a patent or design, you should briefly outline the territories in which you’ll do this.

Publication of details relating to an invention can limit or entirely destroy the potential to patent and commercialise the invention in the future. If you think that patentable Wellcome-funded IP will arise (or when unanticipated IP has arisen), you should explain how you’ll make sure that publications don’t affect your ability to secure and make suitable use of patent protection to advance health benefits.

How IP will be used to achieve health benefits

Guidance:

Wellcome sees IP as a tool which can be used to advance health benefits. You should therefore focus on:

- the benefits your use of the IP will bring to the wider research community
- how this will benefit health.

If your research output is particularly relevant to humanitarian or developing world issues, your plan should specifically address how:

- the output can best be made available for use internationally to address those issues
- your IP strategy will allow this.

Where Wellcome-funded IP comprises a patentable invention, we expect in most cases that it will be protected by filing a patent application. This should be done at a time which maximises the prospects of achieving the desired health benefits, even if this requires a delay to publication. You should only publish details of a potentially patentable invention (without having first sought patent protection) where:

- a market assessment has been carried out and there is no credible prospect of a patent for that invention being commercialised now or in the near future.
- a deliberate decision not to patent the invention (and not to allow anyone else to patent) has been taken for policy reasons. Publication instead of patenting in this case should clearly benefit the wider research community and support the delivery of health benefits. Discuss this with your institution if you’re unsure. Contact Wellcome for advice before publication if you’re still unsure.

Revenue generation should only be a secondary consideration. The primary driver for any commercialisation must be to advance health benefit, even if your employer may generate revenue from commercialising Wellcome-funded IP.

Provide the name and contact details for the person in your organisation (e.g. Technology Transfer Officer or Business Development executive) who can act as a point of contact for Wellcome in connection with the protection and commercialisation of this IP