ABTA

The ABTA understands the importance of data sharing for the advancement of brain tumor research. In that regard, we are currently *asking* Discovery and ABTA Research Collaboration Grantees to state in their progress reports whether they are sharing their data and what repository they are using. Our next step is to incorporate the data sharing plan as a *requirement* in the narrative of our Discovery Grant and ABTA Research Collaboration Grant RFAs next cycle (summer 2018). This will ensure that grantees are thinking carefully about the data they will be generating from the ABTA supported project, and how the data will be used and openly shared. In addition, it will give reviewers an opportunity to review the data sharing plans and offer feedback through their critiques.

St. Baldrick's

St. Baldrick's Data and Resource Sharing Implementation Plan
To reach the most impactful resource and data sharing policies, the St. Baldrick's Foundation is actively working with resource and data sharing experts, scientific advisors, the NCI, and other foundations to understand best practices for both resource and data sharing requirements and reporting the outcomes of resource and data sharing.
This year the following steps are in place:

- Applications: Researchers seeking funds from St. Baldrick's are now required to state their plans for sharing resources and data as part of their application. We are collecting this information for the coming year from applicants and grantees, so that we will be able to assess where our researchers are sharing data and how effective this data sharing is, for both the researcher *sharing* the data and those who *access* it.
- Outcomes reviews: The peer reviewers who conduct outcomes assessment for each St. Baldrick's grant will also specifically address how well the resource and data sharing plans were carried out.

Every two weeks, St. Baldrick's staff meets with big data sharing experts to enable us to better understand big data and data sharing issues. Our scientific advisors are also taking a bigger interest in these issues and offering their guidance. We anticipate this will continue for the coming year, so they can help fine tune our guidelines and policies as the new plan evolves.
We have added the following question to our review process: *Please note the resource sharing plan and if you have additional suggestions related, please include. No score is required for this section.*
Next steps:

1. The following statement appears in our grant guidelines and may be fine-tuned over the coming year as we learn how researchers are responding:
   "*St. Baldrick's Foundation funds biomedical research in order to better understand the causes of pediatric cancers and to advance its prevention, treatment, and cure. The main output of this research is new knowledge. To ensure this knowledge can be accessed, read, applied, and built upon in fulfillment of our goals, the St. Baldrick's Foundation expects its researchers to publish their findings,* ***including but not limited to*** *publication in peer-reviewed journals."*

2. Throughout the year, we will collect responses from applicants and seek feedback from peer-reviewers to refine how we collect and evaluate responses to the above question.

3. We will review guidelines from other foundations, including the Simons Foundation checklist related to resource sharing, to determine if we should incorporate any aspects.
4. In future grant cycles, we will continue to improve and refine the above process.
5. Based on what we learn, we anticipate adopting a more directive policy, most likely for the grant cycles beginning in early 2019.

V Foundation

Addendum: Determining Best Practices for Empowering V Scholars to Analyze and Share Big Data

The V Foundation recognizes that increasingly, cancer researchers are routinely generating large data sets from various types of analyses, including high-throughput genomic data and imaging data. The challenge for scientists in this new world is how best to analyze, integrate and share the large data sets they are producing, to accelerate advances in cancer diagnosis and treatment. As a Foundation which funds the cancer research leaders of tomorrow, we want to support our funded scientists in this challenge.
But how best to do that? We have started work on two projects that will help us determine the best practices for incorporating **data science analysis** and **data-sharing** in future grant support.
Project one: Creation of a new grant mechanism to solicit applications for data science grant proposals in several areas:

1. Fund Translational research that analyzes existing data sets (data mining and sharing grant)
2. Fund training grants for MD trained Clinical Scholars who wish to train in data science analysis.
3. Fund supplementary grants (Mission grants) to provide resources to pay for additional analytical expertise on already funded projects.

We are working with a consultant to create appropriate requests for proposals and helping us to identify the specific data science evaluation criteria and reviewer expertise needed to review the grant proposals.
Project two: Proposing the use of a common Data Platform for data sharing within a new Canine Comparative Oncology Consortium that the V Foundation is helping to create. The incorporation of a common data sharing platform to foster collaborative research between premier research centers at both human (NCI designated cancer centers) and Veterinary schools is expected to accelerate translational research to inform BOTH better patient and pet cancer care.
We are engaging with experts in the data science field such as Warren Kibbe and Martin Ferguson (consultant) to understand the current data science landscape. We are working with a group of stake holders (cancer center directors and leaders at veterinary schools) who are working to create a national Canine Comparative Oncology Consortium and are highly motivated to require data sharing from the start. We are moving forward with the Canine Comparative Consortium and are working on finalizing the governing framework of this new organizations and expectations of members.
Incorporating new data science analysis and data-sharing into the V Scholar program. By the end of 2018, we should have a better idea of how best to provide resources for increased data analysis and require data sharing for our V Scholar Program recipients. At a minimum, we could begin to offer supplemental grants to V Scholars whose projects are generating large data sets to increase their analytical ability (i.e. funding outside expertise).

In many cases, since the V Scholar is a junior faculty member who is just beginning to generate the data that will leverage a sustainable research program, they may need a year to see what their data analytical needs are, and request supplemental funding in year two. Based on an analysis of the types of data being generated, the V Foundation could better develop expectations for the type of data sharing platforms that would be most applicable and requiring data sharing of grant recipients in future years.

American Heart Association

Open Science Policy:
Public Access: The AHA requires that all journal articles resulting from AHA funding be made freely available in PubMed Central within 12 months of publication. It will be the responsibility of the author to ensure this occurs.
Open Data: Any research data that is needed for independent verification of research results must be made freely and publicly available in an AHA approved repository within 12 months of the end of the funding period (and any no-cost extension). A list of Award categories exempt from this requirement is available in the Research Award guide and FAQs. Please also see AHA's Open Science Policy:
http://professional.heart.org/professional/Research/FundingOpportunities/Open-Science-Policy-Statements-for-AHA-Funded-Research_UCM_461225_Article.jsp
In addition to the Open Data Policy requirements, the AHA may require that all research data needed for independent verification of this research must also be provided in an AHA-approved format to an AHA-approved data repository within one year after the end of the Award. If AHA imposes such requirement, AHA will notify Awardee within a reasonable time period after the end of the Award.

Here is the link to our website with information about our Open Science policies, our FAQ, our list of AHA-approved repositories, and our sample data plans.

http://professional.heart.org/professional/ResearchPrograms/AwardsPolicies/UCM_461225_Open-Science-Policy-Statements-for-AHA-Funded-Research.jsp

I will be updating the FAQ document to include more information related to our Public Access policy in the next month or so. We are waiting for the HRA Open software to be deployed so we can anticipate frequent questions that might come up. I have an updated FAQ document that is in draft form currently, but the link I provided only shows what is currently on our website.

Our policies went into effect for awards that started in 2015, so many of them are just now finishing and the investigators have 12 months after their award ends to deposit their data in an approved repository, and they have 12 months after they publish to put the publication in PubMed Central. We will begin to solicit confirmation from the awardees that they have completed these steps and ask them to provide a verifiable link so that we can check compliance. We hope to present an estimate of compliance to our volunteers this spring.

We are able to enforce the requirement that awardees have an acceptable data plan or approved opt-out form because we hold payments on their awards if these items are outstanding or require updating by the investigators. For the post-award requirements, our policy just states that their future funding may be affected is they do not comply.

The AHA has also recently announced the availability of our AHA Precision Medicine Platform. This cloud-based data resource is currently focused on precision cardiovascular medicine and certain awards we fund will be required to put their data in this repository. Portions of this repository will not be open (like dbGaP) because some existing cardiovascular and stroke data sets are not consented for public use and may contain sensitive patient information. We expect to include the open portion of this platform on our list of AHA-approved repositories once it is fully functioning.

Confidential HRA Organization (CHO)

T1D: Data and Biosample Sharing Terms

Sub-Grantee Language:

Grantee shall strive to enter an agreement with [list of applicable sub-grantees] with respect to the Project, (the "Sub-Grant Agreements") within forty-five (45) days following execution of this Agreement and shall promptly inform (CHO) if it is not able to do so and provide (X) the cause of the delay and steps to be taken to resolve the cause of the delay. The Sub-Grant Agreements shall include all terms necessary to enable the Grantee to fulfill the terms of this Agreement; the Sub-Grant Agreements shall be subject to the prior written approval of (CHO). Failure by the Grantee to enter the Sub-Grant Agreements in a form satisfactory to (CHO) within one hundred twenty (120) days following execution of this Agreement shall constitute cause for termination by (CHO) under Section [X] of the Agreement. If and to the extent any term in this Agreement conflicts with the terms of the Sub-Grant Agreements, the terms of this Agreement shall govern.

Steering Committee Language:

Grantee agrees to form a steering committee that will be responsible for the coordination and oversight of (CHO)-funded _____ projects and will be the ultimate decision making body of the _____ Consortium. This committee will be comprised of a representative from each of (CHO) and Grantee, and, for each country in which sponsored research is taking place, a representative from one or more _____ funded institutions.

Grantee agrees that the steering committee will have oversight responsibilities for the implementation of this Agreement's data and biosample sharing provisions, found at Section [X]. The day-to-day responsibilities for compliance with those provisions may be delegated to one or more subcommittees. The steering committee, or if delegated the applicable subcommittee, is responsible for ensuring that the data and samples are made publicly available in accordance with this Agreement in a transparent, cost-efficient, productive, and thoughtful manner.

Should (CHO) exercise its rights under either or both of Section [X] or Section [X] [data and biosample rights], the steering committee will use best efforts to assist (CHO) in transferring oversight responsibilities in relation to making the data and/or biosamples publicly available.

The steering committee must remain in existence for three years following the close of the Grant Period or, if earlier, the completion of the transferring of oversight responsibilities described in the previous paragraph.

Grantee's obligations under this Section survive the Grant Period and/or termination of this Agreement.

Data and Sample Sharing Language:

One of (CHO)'s goals in funding the Grant is to have all funded research be available in the manner most conducive to furthering scientific research. This goal is furthered by the Intellectual Property section of this Agreement, found at Section [X]. In addition, Grantee

agrees to conduct all Grant Activities and manage all Intellectual Property (defined in Section [X] of this Agreement) in a manner that ensures open sharing of data and biosamples as described in subsections (1) and (2) below to the extent legally permissible. Grantee's obligations under this Section survive the Grant Period and/or termination of this Agreement.

The requirements of Section [X] of this Agreement include, but not by way of limitation, that sharing of data and biosamples must comply with all applicable Federal, German, European, and Member State, Data-Protection and Privacy-Protection Laws, Rules and Regulations then in effect.

1. Data Sharing Obligations with respect to Funded Developments:

    a. Deidentified data collected through genetic screening will be made available publicly, at six-months intervals, starting no later than twelve months after the collection of the first DNA sample collected as part of that screening. Grantee agrees to provide (CHO) access to a full copy of deidentified screening data upon request by (CHO). Unless (CHO) has exercised its rights under Subsection (1)(d), (CHO) will consult the steering committee prior to publicly disclosing or using the deidentified data.

    b. Deidentified data collected through clinical trials will be made available publicly no later than twelve months after the completion of the clinical trial, defined as submission of the final trial report to competent regulatory authorities. Grantee agrees to provide (CHO) access to a full copy of deidentified trial data upon request by (CHO) following the completion of the trial. Unless (CHO) has exercised its rights under Subsection (1)(d), (CHO) will consult the steering committee prior to publicly disclosing or using the deidentified data.

    c. Data will be maintained by Grantee or by a third-party public repository approved by (CHO) for a period of at least 1 year following the first date on which the data was made publicly available.

    d. Following this 1-year period, public access to the deidentified data may not be removed unless (CHO) is given prior written notice and an opportunity to maintain, or select a third-party to maintain, the deidentified data. Grantee agrees to use best efforts to assist (CHO) in transferring the deidentified data and any related information should (CHO) exercise this option.

2. Biosample Sharing Obligations with respect to Funded Developments:

    a. At the time a biosample is collected as part of a funded clinical trial, an additional duplicate biosample will be collected where permitted by law if doing so is consistent with guidelines to be developed by the steering committee defined in Section [X]. The additional duplicate biosamples will be made publicly available no later than three years after the first biosample is collected.

    b. Each additional duplicate biosample will be stored at a third-party biobank approved by (CHO) for a period of at least one year following the latter of the close of the Grant Period or the first date on which the biosample was made publicly available.

    c. Following this 1-year period, public access to a biosample may not be removed other than as required by law unless (CHO) is given prior written notice of the reason for removal and an opportunity to elect to have the biosample maintained. Grantee agrees to use best efforts to assist (CHO) in transferring any or all of the biosamples and any related information should (CHO) exercise this option.

Reporting Obligations with Respect to Open Sharing of Data and Biosamples:

The following data, communications, and reports will be provided or made available to (CHO):

1. On a semi-annual basis for the term of the Agreement, an internal audit and report of all data and biosamples collected, shared, or requested for sharing will be provided to (CHO).
2. On an annual basis for the term of the Agreement, an internal data security audit will be performed and reported to (CHO).
3. Requested data and communications will be made available to (CHO) or a third-party selected by (CHO) for the purpose of preparing an annual external audit and report of the data and biosamples collected, shared, or requested for sharing during the Grant Period.

Data Sharing Policy

Goal

One of the greatest obstacles to transformative discovery and progress in the treatment of disease is access to robust and expansive collections of well-curated clinical, biological, and behavioral data. The goal of (CHO) data sharing policy is to create an environment where the broadest research community possible can ethically use data from (CHO) funded projects to advance learnings and accelerate findings that can help people.

The (CHO) is developing an environment, and supporting policies, where grantees will produce, store and utilize (share) data. (CHO) funded projects will be designed to achieve meaningful progress on a per project basis, and the projects will also contribute to a growing and well characterized collection of disease-related data that will be made widely available to any effort intended to improve people's lives.

(CHO)-funded projects involved in the collection of data will adopt the use of data collection standards. These standards will include the use of standard data collection instruments when available, the use of designated data collection schema linking questions to collected data elements, and the use of designated controlled vocabularies to map collected data elements to standard concepts. The goal is to lessen the curation burden for a study's data through the use of these standards.

Some data elements may be collected solely for enhancing the ability to share and semantically connect otherwise disparate data. For example, basic genomic information defining ethnicity would connect external research studies with similar genomic data that might inform on the status and/or progression of the disease being studied in ways that are currently not known. Using the objective information of a participant's genomics, or on the genomics of the study participants in aggregate, connections can be made to understand outcomes and morbidities from external studies that collect genomic information across all disease research. Hypotheses proposed for the studied disease can similarly be tested against any other disease phenotypes to guide future research approaches.

Principles

The goal of the data sharing policy can be understood through the guiding principles below:

1. Dedication to Improving the Lives of People with T1D - data supported by)CHO)-funded grants will be available in the community in a timely manner to help as many people as possible and to translate discoveries into life-saving solutions.
2. Protection of Patients Data - (CHO) will ensure all data collected by (CHO)-funded grants provide the appropriate safeguards that will protect the identity and confidentiality of data.
3. Recognition of Attributions - (CHO) will honor and respect our stakeholder's own incentives. We will recognize the attributions of the investigators, and their collaborators.

4. Accountability and Responsibility - (CHO) supported data will be collected, stored, and shared in a well-defined and consistent process, that follows data standards that guarantee the quality of data, appropriate security, and equitable access to the data. All people who develop, share, and use the data need to be responsible stewards of the data.

Data Policy Guidelines

Implementation

Implementation of the (CHO) Data Sharing Policy has four major foci:

- Development of data sharing policies and processes within (CHO) that will provide the framework for efficient and thoughtful data sharing.
- Definition of the (CHO) proposed project and its governance structure for data, and for any samples collected and stored. This includes the explicit definition of the data and samples to be shared by the project, and the usage criteria surrounding said data and samples.
- Definition of the infrastructure and administrative requirements for a supported project.
- Transfer of data from a project-centric location to a public archival and data access system to ensure post-project access to the data and samples.

These elements are outlined in the sections that follow.

Project and Data Governance

Access to project data may be governed by regulations such as those surrounding clinical trial or observational studies information that will create a necessary framework to outline data sharing. For clinical trial data, the data should be made available no later than twelve months after the completion of the clinical trial, defined as submission of the final trial report to competent regulatory authorities. For non-clinical trial data, the data should be made available no later than twelve months after the (CHO) grant period. Project data will be enhanced as necessary with data to provide connections to external datasets within the disease area studied, and with external datasets for other disease areas.

Single site projects

For a single site project, project and data governance are the responsibility of the principle investigator (PI). If the PI uses a third party to store the data this requires approval in advance by (CHO) and all terms and conditions stated for the PI survive and transfer to the third party. The PI must collaborate with (CHO) to ensure that the informed consents used by the project are consistent with (CHO)'s Data Sharing Policies. The PI is responsible for response to external requests for ancillary project data and on any project alterations that these might involve.

Access to project data and samples are reviewed and dispositioned by the PI; this information must be reported to (CHO) on a regular basis and made available for audit if requested. External audits of project data and samples can be performed annually by (CHO). Guidelines for the acceptance or rejection of data/sample access must be drawn from the (CHO) Data Sharing Policy.

The stewardship of the project data can be assumed by the project PI's institutional data infrastructure or any other professionally managed information technology infrastructure (e.g., Cloud service providers). Project data and collected biospecimins should be released for external use within 12 months of its availability within the project. Collected biospecimins should be released through a biobank after the project has utilized the samples for the funded research project.

Multi-site projects

For multi-site projects project and data governance are implemented in a Steering Committee and a Data (and Sample) Access Committee that can represent the needs and requirements of the respective sites.

1. Steering Committee (SC) – focuses on operational and strategic issues through the review and approval of project direction and changes thereto as they arise. Responsible to approve requests for incremental data types (i.e., that would need to be added to the project data model) for project implementation that are driven by external requests for project data (viz data sharing). When possible, responsible for setting the desired scope of the informed consents that are used by the project. It is desirable for a project's Informed Consents to be reviewed and approved by (CHO) for compatibility with the (CHO) Data Sharing Policy. The SC is responsible for oversight of the Data Access Committee (DAC) and any other committee(s) required for successfully managing the project.

   The composition of the SC should have representation by clinical site leaders augmented by a (CHO) representative and a leader from the data coordination center used for the project. The SC should represent all site- and country-specific issues.

a. Ownership of the project data needs to be public (with all patient identifying information removed).

b. Scope of informed consent should explicitly and ethically allow for use of data by health-related research (academic, not-for-profit, and commercial). Samples and data are not sold.

c. Scope of informed consent should acknowledge research use beyond a specific disease. That is, study samples and data can be used for health-related research beyond studied disease to broaden the research community's understanding of conditions that might additionally affect studied patients.

   i. Suggested language is: "Study samples and data will be shared with qualified researchers to perform health-related research. Such sharing will be approved by a DAC, and reviewed by an IRB. Independent researchers may publish their findings but may not otherwise share the data with anyone that has not been so authorized by the DAC."

d. Informed consent should acknowledge research use. Health-related research can be very broad in terms of the objective of the research, including understanding population health relationships, treatment outcomes and adverse events, and use of the study cohort as a control group.

e. Informed consent should include a section on "How will the study data and samples be used?" that will call out the study hypotheses being explored, and that includes a section on the sharing of the samples and data for health-related research to advance our understanding of the study participants potentially beyond our understanding of the studied disease directly. (see 1.b and 1.c above)

2. Data Access Committee (DAC) – reviews and approves requests for access to project data and stored samples that are available for sharing through submitted Data/Sample Access Agreements (DAA and SAA – see below). The DAC is responsible to ensure that requests are consistent with participant informed consent and any patient privacy requirements that may be in existence (e.g., sharing of data between EU and non-EU entities). The DAC should rely on support groups such as an IRB as needed. Membership should reflect (CHO) and project data contributors. Oversight of DAC provided by the Steering Committee. Access decisions guided by agreed embargo limits. For sample access, consideration of each request should be evaluated against sample availability and intended research use being proposed.

a. Data Access Agreement (DAA) for external use of research data is a contract between the Data Coordination Center (on behalf of the project) and the Data User that includes the details of data use, publication embargoes, data attribution, and data storage (i.e., once the data is downloaded). DAAs should call out named individuals rather than organizations or research groups. See separate attachment drafted from an EMBL/EBI DAA.

b. A Sample Access Agreement (SAA) for external use of stored biospecimins is a contract between the biobank and the Researcher on behalf of the project. SAAs should call out named individuals and their organizations. The SAA should be in place once samples have been collected and submitted to the project's biospecimin repository.

3. Data Coordination Center (DCC) – serves as the steward of project data. Responsible for maintaining the security of the project data as well as a log of data access that can be audited. The DCC reports into the project Steering Committee. Project data should be released for external use within 12 months of its availability within the project.
4. Biospecimin Repository (biobank) – serves as the steward of the project's biospecimins. Responsible for maintaining the integrity of the samples, and for maintaining an access log for audit purposes. The biobank reports into the Steering Committee. Collected biospecimins should be released through a biobank after the project has utilized the samples for the funded research project, and within 12 months of their availability within the project.
5. Governance of the project by (CHO) will be through the project's Steering Committee. The SC will internally audit data and biosamples that are stored/collected, shared, and requested and not shared every six months for the duration of the grant and submit a report for review to (CHO). (CHO) will externally audit the same information on an annual basis.

Infrastructure and Administration

The data location(s), data mobility, and technical administration of the data must be defined explicitly. Certifications for data security and disaster recovery should be in place with their attendant audit schedules.

Single site projects

Single site projects are encouraged to use their institutional data center or any other professionally managed information technology infrastructure (e.g., Cloud service providers) to store their data. A data administrator must be named in the application who can certify the integrity of the data stored. If local computation on the data is required, due to data use or regulatory constraints, it should be provided through collaboration with the PI. A post-project data sustainment plan should be proposed by the PI.

Multi-site projects

For larger projects with larger data sets there may be a need to provide compute resources so that Data Users can deploy compute to the data center (e.g., using Docker or similar technologies) so that data does not need to move outside of the data center(s). If required (by data use requirements and/or regulations), the need for compute and embedding technology will drive the need for additional hardware and personnel resources within the project that should be called out in the grant proposal.

1. Each project requires a Data Coordination Center to be named that will accept the responsibilities incumbent for stewardship of the project's data. A sustainment plan for the data should be in place for when the project completes and the project is no longer receiving funding.
2. The data security and disaster recovery plans for the data, and who has administrative access to the data, will be reviewed.

3. A training plan for the clinical (data) platform must be established that includes documentation and outreach to external groups that can benefit from the platform.

Transfer of Data from Project to Public Archival System

The sustainment of project data beyond the duration of the funded research is an important aspect of the (CHO) Data Sharing Policy. In addition to the need to document the acceptable use of the data (eg, if there are restrictions within the informed consents of the participants), it is also important to provide adequate metadata that fully documents the data elements and their inter-relationships. Adequate metadata is further defined as the necessary descriptive content to permit the loading of the project data into a public archival system.

Single site projects

All project data should be aggregated throughout the project, and must be complete before the project ends. On an annual basis, a backup copy of the data should be transferred to a secure site designated by (CHO). All data fields and data schema must be fully documented through metadata.

Multi-site projects

The aggregation of all project data from each site and each PI to the data coordination center should be achieved throughout the funded project, and must be complete before the project ends. Similarly, on an annual basis the aggregated project data from the data coordination center should be transferred to the (CHO) program manager as a backup. Ideally this backup copy will be created as part of the annual external data audits outlined above. As noted previously, all data fields and data schema should be fully documented through metadata.

Potential Policy Exceptions

There may be several possible exceptions to the (CHO) Data Sharing Policy that are helpful to document. For example, a project might be proposed that accesses historic samples that were collected with informed consents that are more restrictive than the (CHO) policy. Another is where the amount of data collected is small enough to render data sharing un-useful. An another is where the shortness of the project's duration, perhaps as a pilot study prior to a larger project, argues against the added administration that the Data Sharing Policy requires.

Weston Brain Institute

5.2. The Institute expects results of funded research to be published as rapidly as possible in the open access scientific literature or other forms of publication that are readily available to the general public and/or research community. Such publication should be consistent with high standards of scientific excellence and rigor, and provide sufficient detail so that the research community can benefit from the findings from or in connection with the Project. You agree to advise the Institute of the details of any such publication (journal, publication date, etc.), as soon as the information is known.

A lay person abstract of the research proposal must be submitted prior to funding. A lay person abstract of the research results must be submitted no later than 1 month from the date of grant expiration. These abstracts may be made available to the public by the Institute.

5.3. The Institute expects that all tools or reagents funded by it or resulting from the funded projects, should be made readily available to the research community either freely or at reasonable prices within 9 months of study completion. If sharing of such tools or reagents will jeopardize the Applicant's right to secure patents or copyrights necessary to protect the Applicant's ownership, then they should be made available as soon these rights have been secured. The Institute may let the public know of these tools or reagents so other researchers know they are available.

5.4. The Institute requires that any clinical trial granted under any of its funding programs be registered with clinicaltrials.gov, PDtrials.org or other appropriate public registry.

American Diabetes Association

- Research Programs Data and Resource Sharing Policy
-
- Research resources and data derived through American Diabetes Association (ADA) sponsored research are critical to the advancement of scientific progress. Results and resources developed with ADA funds are required to be made available to the broader scientific community within a reasonable timeframe, as defined within this policy. Applications for funding must include a data and resource sharing plan, or a request for waiver. The plan or the waiver will be evaluated by the ADA's Research Grant Review Committee. Data and Resource Sharing Plan (Required in all grant applications, effective January 1, 2018) Open Data Sharing All data resulting from ADA-funded research that can be shared without compromising human subject protections must be shared to an approved open data repository within 6 months of publication or within 18 months of the conclusion of the funding period, if the study remains unpublished. A list of vetted, free repositories will be maintained at professional.diabetes.org/research-grants. This list includes repositories recommended by NIH. Awardees are encouraged to use the repository most appropriate for the subject matter of the research conducted. If an awardee desires to use an unlisted repository, the ADA will review the appropriateness of the requested repository.
-
- To be approved, the new repository must be:

    - Secure: Employs a satisfactory security policy to ensure that datasets are stored and that confidential information is protected.
    - Stable: Provides reasonable assurances that the repository will be maintained and accessible indefinitely.
    - Free to Access and Use: Allows interested parties to access and use the data for research purposes without restriction (except where human subject grounds are applicable).
    - Provide Searchable Metadata: Data are searchable for ease of access and use.
    - Allow Applicable File Formats: Accommodates all appropriate file types.

Resource Sharing Resources developed with ADA grant funding are required to be made available to the broader scientific community. In particular, ADA-funded projects expected to generate unique model organism resources or genomic data must include specific plans for sharing and distributing. If sharing is not possible, the application must include an acceptable explanation and request for waiver. In general, to the extent possible, ADA grantees are expected to share all scientific resources upon request for the advancement of

research progress. While the data and resource sharing plan will not impact the application score, it is a requirement for submission. A request for waiver may be made in one of the following categories:

- Human Subject Protection (privacy regulations or consent of research participants)
- Superseding Regulations (laws or institutional policies)
- Intellectual Property (existing IP rights)

If a request for waiver is not approved and the grant is funded, the applicant will be required to submit a data and resource sharing plan. Enforcement All grantees will be required to affirm adherence to their data and resource sharing plan in accordance with this policy. At final reporting or within 18 months of the grant term, grantees will be required to affirm that data has been uploaded to an approved repository. Grantees who fail to comply with this policy may not be eligible for subsequent grant funding from the ADA. Furthermore, at the discretion of the Research Policy Committee, grantees who fail to comply with this policy may be listed in a publication of research misconduct and ethics violations.

Children's Tumor Foundation

- This apply only to the SYNODOS grants. All other grants do not require data sharing, we only encourage them to publish and share the results, but no requirements. These are the points as written in the synodos award contracts with all the participating researchers. 7.1 Except as permitted in the Funding Conditions, none of the Parties will ,during the Confidentiality Period, disclose to any third party nor use for any purpose, except carrying out the Project or as otherwise permitted by this Agreement or by the Funding Conditions, any other Party's other Confidential Information. The Confidentiality Period is 7.1.1. set to three (3) years after disclosure for any Background and 7.1.2 set out in the Funding Conditions for any portion of Results. THEN THE FUNDING CONDITIONS SAY THE FOLLOWING: 1. Periodic Reports and data sharing 1.1 The Periodic Reports will be accompanied upon request from the Project Coordinator, by the submission of all data in an electronic format suitable for insertion into a repository of CTF's choice. 1.2 During the Confidentiality Period, access to this data in this repository will be limited to CTF and any Funded Party that has not withdrawn or is deemed to have withdrawn pursuant to clauses 10.1 and 10.2 of the Agreement 1.3 At the end of the Confidentiality Period, CTF shall give access to this data in this repository to anyone without restriction. (…) 2.4 During the Confidentiality Period, CTF shall have the right to use the Results in order to make decisions as to whether or not to fund other projects or grants. (…) 3. Confidentiality 3.1 The confidentiality period of clause 7.1.2 of the Agreement shall be decided by CTF on a case by case basis , such decision to take into account the relevant Funded Party's arguments. 3.2 Unless otherwise agreed by CTF and the relevant Funded Parties, in no event shall this confidentiality period be less than six (6) months from the disclosure within the Periodic Report 3.3 In no event shall this confidentiality period exceed one (1) calendar year from the disclosure within the Periodic Report

BRAIN Commons

1. **Resource Sharing Plan**: Individuals are required to comply with the instructions for the Resource Sharing Plans as provided in the SF424 (R&R) Application Guide, with the following modification:
    - All applications, regardless of the amount of direct costs requested for any one year, should address a Data Sharing Plan.

   This guidance provides the National Institutes of Health (NIH) policy statement on data sharing and additional information on the implementation of this policy.

- Goals of Data Sharing
- Applicability

GOALS OF DATA SHARING

Data sharing promotes many goals of the NIH research endeavor. It is particularly important for unique datathat cannot be readily replicated. Data sharing allows scientists to expedite the translation of research results into knowledge, products, and procedures to improve human health.

There are many reasons to share data from NIH-supported studies. Sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new datasets when data from multiple sources are combined.

In NIH's view, all data should be considered for data sharing. **Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data.** To facilitate data sharing, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.

APPLICABILITY

The NIH policy on data sharing applies:

- To the sharing of final research data for research purposes.
- To basic research, clinical studies, surveys, and other types of research supported by NIH. It applies to research that involves human subjects and laboratory research that does not involve human subjects. It is especially important to share unique data that cannot be readily replicated.
- To applicants seeking $500,000 or more in direct costs in any year of the proposed project period through grants, cooperative agreements, or contracts.
- To research applications submitted beginning October 1, 2003.

Policies with respect to data sharing vary across countries. Investigators from foreign institutions and U.S. investigators collecting data in other countries should

familiarize themselves with the policies governing data sharing in the countries in which they plan to work and to address any specific limitations in the data-sharing plan in their application.

Even if NIH support is sought to transform or link datasets (as opposed to producing a new set of data), the investigator should still include a data-sharing plan in the application. If there are limitations associated with a data-sharing agreement for the original data that preclude subsequent sharing, then the applicant should explain this in the application.

IMPLEMENTATION

The NIH data-sharing policy applies to applicants seeking $500,000 or more in direct costs in any year of the proposed research. The $500,000 threshold corresponds to the threshold set in the October 16, 2001 NIH Guide, where applicants requesting $500,000 or more in direct costs for any year must seek agreement by NIH Institute or Center (IC) staff to accept assignment of their application at least 6 weeks prior to the anticipated submission date. (See http://grants2.nih.gov/grants/guide/notice-files/NOT-OD-02-004.html). That policy directs applicants to contact in writing or by telephone IC program staff during the development process of the application but no later than 6 weeks before the anticipated submission date. Applicants are encouraged to discuss their proposed data-sharing plan with IC program staff at that time.

Final research data are recorded factual material commonly accepted in the scientific community as necessary to document, support, and validate research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based. For most studies, final research data will be a computerized dataset. For example, the final research data for a clinical study would include the computerized dataset upon which the accepted publication was based, not the underlying pathology reports and other clinical source documents. For some but not all scientific areas, the final dataset might include both raw data and derived variables, which would be described in the documentation associated with the dataset.

Given the breadth and variety of science that NIH supports, neither the precise content for the data documentation, nor the formatting, presentation, or transport mode for data is stipulated. What is sensible in one field or one study may not work at all for others. It would be helpful for members of multiple disciplines and their professional societies to discuss data sharing, determine what standards and best practices should be proposed, and create a social environment that supports data sharing. NIH is planning to convene workshops where investigators with experience in data sharing will share their expertise with others. These workshops will address areas such as cleaning and formatting data, writing documentation, redacting data to protect subjects' identities and proprietary information, and estimating costs to prepare documentation and data for sharing.

When the Principal Investigator (PI) and the authorized institutional official sign the face page of an NIH application, they are assuring compliance with policies and regulations governing research awards. NIH expects grantees to follow these rules and to conduct the work described in the application. Thus, if an application describes a data-sharing plan, **NIH expects that plan to be enacted**. If progress has been made with the data-sharing plan, then the grantee should note this in the progress report. In the final progress report, if not sooner, the grantee should note what steps have been taken with respect to the data-sharing plan. In the case of noncompliance (depending on its severity and duration) NIH can take various

actions to protect the Federal Government's interests. In some instances, for example, NIH may make data sharing an explicit term and condition of subsequent awards.

Grantees should note that, under the NIH Grants Policy Statement, they are required to keep the data for 3 years following closeout of a grant or contract agreement. (Contracts may specify different time periods.) For the most part, NIH makes awards to institutions and not individuals (with very few exceptions, such as F32 awards). Thus, the grantee institution may have additional policies and procedures regarding the custody, distribution, and required retention period for data produced under research awards.

Timeliness of Data Sharing

Recognizing that the value of data often depends on their timeliness, data sharing should occur in a timely fashion. NIH expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. The specific time will be influenced by the nature of the data collected. Data from small studies can be analyzed and submitted for publication relatively quickly. If data from large epidemiologic or longitudinal studies are collected over several discrete time periods or waves, it is reasonable to expect that the data would be released in waves as data become available or main findings from waves of the data are published. NIH recognizes that the investigators who collected the data have a legitimate interest in benefiting from their investment of time and effort. NIH continues to expect that the initial investigators may benefit from first and continuing use but not from prolonged exclusive use.

Human Subjects and Privacy Issues

The rights and privacy of human subjects who participate in NIH-sponsored research must be protected at all times. It is the responsibility of the investigators, their Institutional Review Board (IRB), and their institution to protect the rights of subjects and the confidentiality of the data. Prior to sharing, data should be redacted to strip all identifiers, and effective strategies should be adopted to minimize risks of unauthorized disclosure of personal identifiers. Stripping a dataset of items that could identify individual participants is referred to by several different terms, such as "data redaction," "de-identification of data," and anonymizing data. In addition to removing direct identifiers, e.g., name, address, telephone numbers, and Social Security Numbers, researchers should consider removing indirect identifiers and other information that could lead to "deductive disclosure" of participants' identities. Deductive disclosure of individual subjects becomes more likely when there are unusual characteristics of the joint occurrence of several unusual variables. Samples drawn from small geographic areas, rare populations, and linked datasets can present particular challenges to the protection of subjects' identities.

Investigators may use different methods to reduce the risk of subject identification. One possible approach is to withhold some part of the data. Another approach is to statistically alter the data in ways that will not compromise secondary analyses but will protect individual subjects' identities. Alternatively, an investigator may restrict access to the data at a controlled site, sometimes referred to as a data enclave. Some investigators may employ hybrid methods, such as releasing a highly redacted dataset for general use but providing access to more sensitive data with stricter controls through a data enclave.

Researchers who seek access to individual level data are typically required to enter into a data-sharing agreement. Data-sharing agreements, which come by many

terms, including "license agreements," and "data distribution agreements," generally include requirements to protect participants' privacy and data confidentiality. They may prohibit the recipient from transferring the data to other users or require that the data be used for research purposes only, among other provisions, and they may stipulate penalties for violations. For further information on these alternative mechanisms to share data while protecting participant confidentiality, see also the section concerning "Methods for Data Sharing." In most instances, sharing and archiving of data is possible without compromising confidentiality and privacy rights. The procedures adopted to share data while protecting privacy should be individually tailored to the specific dataset.

Investigators seeking NIH support for clinical trials may wish to consider several factors as they develop their data-sharing plan. Researchers who are planning clinical trials and intend to share the resulting data should think carefully about the study design, the informed consent documents, and the structure of the resulting dataset prior to the initiation of the study. For example, many early phase clinical trials use small samples, which make it difficult to protect the privacy of the participants. Furthermore, some study designs afford greater privacy protection to subjects than others. For example, longitudinal research poses challenges because the need to retain identifiers in order to link individual-specific data collected at different time points.

NIH recognizes that the sharing of data from clinical trials and under other situations may require making the data anonymous or sharing under more controlled means, as through a restricted access data enclave. Sharing though data enclaves would grant access only to researchers who agree to preserve the privacy of subjects and provide means to protect the confidentiality of the data.

Investigators who are working for or who are themselves covered entities under the Health Insurance Portability and Accountability Act (HIPAA) must consider issues related to the **Privacy Rule**, a Federal regulation under HIPAA that governs the protection of individually identifiable health information. The Department of Health and Human Services (DHHS) provides guidance on research and the Privacy Rule elsewhere (http://www.hhs.gov/ocr/). It should be noted that the Privacy Rule is relatively new, and additional information and guidance will be shared on the DHHS website as soon as it is available.

If research participants are promised that their data will not be shared with other researchers, the application should explain the reasons for such promises. Such promises should not be made routinely and without adequate justification. For the most part, it is not appropriate for the initial investigator to place limits on the research questions or methods other investigators might pursue with the data. It is also not appropriate for the investigator who produced the data to require coauthorship as a condition for sharing the data.

Many research efforts supported by NIH do not include human subjects. Final research datasets from studies that do not include human subjects generally should not be constrained by the limitations deemed necessary and appropriate for human subjects.

Proprietary Data

Although Small Business Innovation Research (SBIR) applicants are also to address data sharing in their applications, under the Small Business Act, SBIR grantees may withhold their data for 4 years after the end of the award. The Small Business Act provides authority for NIH to protect from disclosure and nongovernmental use all SBIR data developed from work performed under an SBIR funding agreement for a

period of 4 years after the closeout of either a phase I or phase II grant unless NIH obtains permission from the awardee to disclose these data. The data rights protection period lapses only upon expiration of the protection period applicable to the SBIR award, or by agreement between the small business concern and NIH. Issues related to proprietary data also can arise when cofunding is provided by the private sector (e.g., the pharmaceutical or biotechnology industries) with corresponding constraints on public disclosure. NIH recognizes the need to protect patentable and other proprietary data. Any restrictions on data sharing due to cofunding arrangements should be discussed in the data-sharing plan section of an application and will be considered by program staff. While NIH understands that an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 30 to 60 days is generally viewed as a reasonable period for such activity.

Methods for Data Sharing

There are many ways to share data.

- Under the auspices of the PI
- Data archive
- Data enclave
- Mixed mode sharing.

The method for sharing that an investigator selects is likely to depend on several factors, including the sensitivity of the data, the size and complexity of the dataset, and the volume of requests anticipated. Investigators sharing under their own auspices may simply mail a CD with the data to the requestor, or post the data on their institutional or personal Website. Although not a condition for data access, some investigators sharing under their own auspices may form collaborations with other investigators seeking their data in order to pursue research of mutual interest. Others may simply share the data by transferring them to a data archive facility to distribute more widely to interested users, to maintain associated documentation, and to meet reporting requirements. Data archives can be particularly attractive for investigators concerned about a large volume of requests, vetting frivolous or inappropriate requests, or providing technical assistance for users seeking help with analyses.

There are several mechanisms for data sharing that investigators can use. For example, investigators sharing under their own auspices should consider using a **data-sharing agreement** to impose appropriate limitations on users. Such an agreement usually indicates the criteria for data access, whether or not there are any conditions for research use, and can incorporate privacy and confidentiality standards to ensure data security at the recipient site and prohibit manipulation of data for the purposes of identifying subjects. Many examples of data sharing agreements for specific datasets are available on the Internet, including the following:

AHRQ National Inpatient Sample at
http://www.ahcpr.gov/data/hcup/datause.htm

Russian Longitudinal Monitoring Survey at
http://www.cpc.unc.edu/dataarch/iprimary/rlms.html

Center for Medicare and Medicaid Services Data at
http://hrsonline.isr.umich.edu/rda/userdocs/cmsdua.pdf (PDF - 59 KB)

Alternatively, researchers may want to add their data to a data archive or a data enclave. Datasets that cannot be distributed to the general public, for example,

because of participant confidentiality concerns, third-party licensing or use agreements that prohibit redistribution, or national security considerations, can be accessed through a data enclave. A data enclave provides a controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.

Investigators may also wish to develop a "mixed mode" for data sharing that allows for more than one version of the dataset and provides different levels of access depending on the version. For example, a redacted dataset could be made available for general use, but stricter controls through a data enclave would be applied if access to more sensitive data were required.

Investigators will need to determine which method of data sharing is best for their particular dataset. The Data Sharing Workbook (PDF - 75 KB) or (MS Word - 74 KB) provides information and examples of how others have shared data.

Data Documentation

Regardless of the mechanism used to share data, each dataset will require documentation. (Some fields refer to data documentation by other terms, such as metadata or codebooks). Proper documentation is needed to ensure that others can use the dataset and to prevent misuse, misinterpretation, and confusion. Documentation provides information about the methodology and procedures used to collect the data, details about codes, definitions of variables, variable field locations, frequencies, and the like. The precise content of documentation will vary by scientific area, study design, the type of data collected, and characteristics of the dataset.

It is appropriate for scientific authors to **acknowledge the source of data** upon which their manuscript is based. Many investigators include this information in the methods and/or reference sections of their manuscripts. Journals generally include an acknowledgement section, in which the authors can recognize people who helped them gain access to the data. Authors using shared data should check the policies of the journal to which they plan to submit to determine the precise location in the manuscript for such acknowledgement. Most journals now expect that DNA and amino acid sequences that appear in articles will be submitted to a sequence database before publication.

Funds for Data Sharing

NIH recognizes that it takes time and money to prepare data for sharing. Thus, applicants can request funds for data sharing and archiving in their grant application. (See also the section on What to Include in an NIH Application.)

Investigators who incorporate data sharing in the initial design of the study may more readily and economically establish adequate procedures for protecting the identities of participants and share a useful dataset with appropriate documentation.

Review Considerations

Reviewers will not factor the proposed data-sharing plan into the determination of scientific merit or priority score. Program staff will be responsible for overseeing the data sharing policy and for assessing the appropriateness and adequacy of the proposed data-sharing plan.

WHAT TO INCLUDE IN AN NIH APPLICATION

Investigators seeking $500,000 or more in direct costs in any year should include a description of how final research data will be shared, or explain why data sharing is not possible. It is expected that the data sharing discussion will be provided primarily in the form of a brief paragraph immediately following the Research Plan

Section of the PHS 398 application form (i.e., immediately after I. Letters of Support), and **would not count towards the application page limit**.

Data Sharing Plan (to follow immediately after the Research Plan Section)

The precise content of the data-sharing plan will vary, depending on the data being collected and how the investigator is planning to share the data. Applicants who are planning to share data may wish to describe briefly the expected schedule for data sharing, the format of the final dataset, the documentation to be provided, whether or not any analytic tools also will be provided, whether or not a data-sharing agreement will be required and, if so, a brief description of such an agreement (including the criteria for deciding who can receive the data and whether or not any conditions will be placed on their use), and the mode of data sharing (e.g., under their own auspices by mailing a disk or posting data on their institutional or personal website, through a data archive or enclave). Investigators choosing to share under their own auspices may wish to enter into a data-sharing agreement. References to data sharing may also be appropriate in other sections of the application, as discussed below.

Budget and Budget Justification Sections

Applicants may request funds in their application for data sharing. If funds are being sought, the applicant should address the financial issues in the budget and budget justification sections. Some investigators have more experience than others in estimating costs associated with preparing the dataset and associated documentation, and providing support to data users. As investigators gain experience with the process, their ability to estimate costs will improve. Investigators working with archives can get help with data preparation and cost estimation. Investigators who are concerned about paying for data-sharing costs at the end of their grant can make prior arrangements with archives. Investigators facing considerable delays in the preparation of the final dataset for sharing should consult with the NIH program about how to manage this situation, such as requesting a no-cost extension.

Background and Significance Section (PHS 398 Research Plan Section B)

If support is being sought to develop a large database that will serve as an important resource for the scientific community, the applicant may wish to make a statement about this in the significance section of the application.

Human Subjects Section (PHS 398 Research Plan Section E)

If the research involves human subjects and the data are intended to be shared, the application should discuss how the rights and confidentiality of participants would be protected. In the Human Subjects section of the application, the applicant should discuss the potential risks to research participants posed by data sharing and steps taken to address those risks.

EXAMPLES OF DATA-SHARING PLANS

The precise content and level of detail to be included in a data-sharing plan depends on several factors, such as whether or not the investigator is planning to share data, the size and complexity of the dataset, and the like. Below are several examples of data-sharing plans.

Example 1

The proposed research will involve a small sample (less than 20 subjects) recruited from clinical facilities in the New York City area with Williams syndrome. This rare craniofacial disorder is associated with distinguishing facial features, as well as mental retardation. Even with the removal of all identifiers, we believe that it would be difficult if not impossible to protect the identities of subjects given the

physical characteristics of subjects, the type of clinical data (including imaging) that we will be collecting, and the relatively restricted area from which we are recruiting subjects. Therefore, we are not planning to share the data.

Example 2

The proposed research will include data from approximately 500 subjects being screened for three bacterial sexually transmitted diseases (STDs) at an inner city STD clinic. The final dataset will include self-reported demographic and behavioral data from interviews with the subjects and laboratory data from urine specimens provided. Because the STDs being studied are reportable diseases, we will be collecting identifying information. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics. Thus, we will make the data and associated documentation available to users only under a data-sharing agreement that provides for: (1) a commitment to using the data only for research purposes and not to identify any individual participant; (2) a commitment to securing the data using appropriate computer technology; and (3) a commitment to destroying or returning the data after analyses are completed.

Example 3

This application requests support to collect public-use data from a survey of more than 22,000 Americans over the age of 50 every 2 years. Data products from this study will be made available without cost to researchers and analysts.

https://ssl.isr.umich.edu/hrs/

User registration is required in order to access or download files. As part of the registration process, users must agree to the conditions of use governing access to the public release data, including restrictions against attempting to identify study participants, destruction of the data after analyses are completed, reporting responsibilities, restrictions on redistribution of the data to third parties, and proper acknowledgement of the data resource. Registered users will receive user support, as well as information related to errors in the data, future releases, workshops, and publication lists. The information provided to users will *not* be used for commercial purposes, and will *not* be redistributed to third parties.

DEFINITIONS

Covered Entity - A covered entity is defined as a health care clearinghouse, health plan, or health care provider that electronically transmits health information in connection with a transaction for which DHHS has adopted standards under the Health Insurance Portability and Accountability Act (HIPAA). An example of a researcher who may be a covered entity is a physician who electronically bills for health care services and conducts clinical trials. A set of decision tools on "Am I a covered entity?" are available from the DHHS Office for Civil Rights Website http://www.hhs.gov/ocr/

Data - see Final Research Data

Data Archive - A place where machine-readable data are acquired, manipulated, documented, and finally distributed to the scientific community for further analysis.

Data Enclave - A controlled, secure environment in which eligible researchers can perform analyses using restricted data resources.

Final Research Data - Recorded factual material commonly accepted in the scientific community as necessary to document and support research findings. This does not mean summary statistics or tables; rather, it means the data on which summary statistics and tables are based. For the purposes of this policy, final

research data do not include laboratory notebooks, partial datasets, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as gels or laboratory specimens. NIH has separate guidance on the sharing of research resources, which can be found at https://grants.nih.gov/grants/policy/nihgps_2013/nihgps_ch8.htm#_Toc271264947

Restricted Data - datasets that cannot be distributed to the general public, because of, for example, participant confidentiality concerns, third-party licensing or use agreements, or national security considerations.

Timeliness - In general, NIH considers the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. However, the actual time will be influenced by the nature of the data collected.

Unique Data - Data that cannot be readily replicated. Examples of studies producing unique data include: large surveys that are too expensive to replicate; studies of unique populations, such as centenarians; studies conducted at unique times, such as a natural disaster; studies of rare phenomena, such as rare metabolic diseases.

Cancer Research UK

Data sharing guidelines

At CRUK, we are committed to ensuring that the data generated through its funding should be put to maximum use by the cancer research community and, whenever possible, is translated to deliver patient benefit. It is therefore our policy that all data generated as a result of our funding be considered for sharing and made as widely and freely accessible as possible whilst safeguarding intellectual property, the privacy of patients and confidential data. Researchers applying for funding should familiarise themselves with our Data Sharing and Preservation Policy. Given the diverse nature of the research we support, the guidelines below do not prescribe precisely how and when investigators should share research data. Instead they should be used to ensure that the principles of the policy are adhered to.

Applicability

Our Data Sharing and Preservation Policy is applicable to all candidates seeking CRUK funding after 1 April 2009 and applies:

- To the sharing of final research data for research purposes.
- To basic research, clinical studies, surveys and other types of research supported by CRUK.
- Especially to unique data that cannot readily be replicated.
- To projects that transform or link pre-existing datasets.

The data from all activities in the preparation for and arising out of phase 1 and 2 clinical trials which CRUK sponsors and which are initiated by its Centre for Drug Development, after approval by its New Agents Committee, is not automatically covered under this Data Sharing Policy. For clarity on the position, please contact the Centre for Drug Development on a trial by trial basis.

Data management and sharing plan

All applicants seeking funding from CRUK will be required to submit a data sharing plan as part of their research grant proposal. If data sharing is not appropriate, applicants must include a clear explanation why. The data sharing plan will be reviewed as part of the funding decision. Funding committees will assess the appropriateness and adequacy of the data sharing plan and provide specific feedback to applicants where necessary.

We recognise that data sharing strategies will vary according to the type of data collected and thus do not specify the exact content and format of the data sharing plan. We recommend that data should be shared using established standards and existing resources where possible. The following should be considered when developing a data sharing plan:

- The volume, type, content and format of the final dataset
- The standards that will be utilised for data collection and management
- The metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly
- The method used to share data
- The timescale for public release of data
- The long-term preservation plan for the dataset Whether a data sharing agreement will be required
- Any reasons why there may be restrictions on data sharing, for example;
- Development arrangements with our Commercial Partnerships team including intellectual property protection and commercialisation
- Proprietary data – restrictions due to collaborations with for profit organisations International policies governing the sharing of data collected outside of the UK
- Confidentiality, ethical or consent issues that may arise with the use of data involving human subjects.

Funding committees will monitor investigators' progress in implementing their data management and sharing plan. However, we understand that an investigator may need to adapt the method and timelines for sharing during the course of the study – for example, when potential intellectual property arises unexpectedly.

Intellectual property rights and proprietary data

Data which might have the potential to be exploited commercially or otherwise to deliver patient benefit should be discussed with your technology transfer office and ourCommercial Partnerships team prior to data sharing.

We encourage the appropriate filing of patents and recognises that there may be a need to delay the release of data until patent applications have been filed. Whilst there may be a delay in the release of data due to the application process, appropriate intellectual property protection should not hinder data sharing and may be the best way of ensuring that patient (and public) benefit is delivered.

Any intellectual property issues or plans for commercialisation that may affect data sharing should be addressed in the data sharing plan. We understand that unexpected intellectual property may arise during the course of the study and investigators may need to depart from their data sharing plan to protect intellectual property and for any other necessary steps to be taken.

Data sharing may also be affected when co-funding is provided by the private sector (e.g. by a pharmaceutical company) or host institution resulting in some restrictions on the disclosure of data. For example with clinical trials, the Trial Management Group and/or trial sponsor etc may impose restrictions on data access. Any restrictions should be outlined in the data sharing plan and applicants should explore ways data sharing requests can be considered by the body that owns the data.

Standards, metadata and documentation

For data sharing to be a success it is important that data are prepared in such a way that those using the dataset have a clear understanding of what the data mean so that they can be used appropriately. To enable this, applicants are encouraged to include with the dataset all the necessary information (metadata) describing the data and their format. This information should include such information as the methodology used to collect data, definitions of variables, units of measurement, any assumptions made, the format of the data, file type of the data etc. To support this researchers are strongly encouraged to utilise community standards to describe and structure data, (e.g. common terminology, minimum information guidelines and standard data exchange formats).

Methods for data sharing

The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Data can be shared by any of the following methods:

Under the auspices of the Principal Investigator

Investigators sharing under their own auspices may securely send data to a requestor, or upload the data to their institutional website. Investigators should consider using a data-sharing agreement (see below) to impose appropriate limitations on the secondary use of the data.

Through a third party

Investigators can share their data by transferring it to a data archive facility to distribute more widely to the scientific community, to maintain documentation and meet reporting requirements. Data archives are particularly attractive for investigators concerned about managing a large volume of requests for data, vetting frivolous or inappropriate requests, or providing technical assistance for users seeking to help with analyses.

Using a data enclave

Datasets that cannot be distributed to the general public due to confidentially concerns, or third-party licensing or use agreements that prohibit redistribution, can be accessed through a data enclave. A data enclave provides a controlled secure environment in which eligible researchers can perform analyses using restricted data resources.

Through a combination of methods

Investigators may wish to share their data by a combination of the above methods or in different versions, in order to control the level of access permitted.

Timeframe for data sharing

As the value of data is often dependent on its timeliness, we expect that data sharing should occur in a timely manner. We acknowledge that the investigators who generated the data have a legitimate interest in benefiting from their investment of time and effort and we therefore support the initial investigator having a reasonable period of private use of the data but not prolonged exclusive use.

We expect data to be released no later than the acceptance for publication of the main findings from the final dataset (unless restrictions from third party agreements or IP protection still apply) or on a timescale in line with the procedures of the relevant research area. For example, for crystallography data there is an agreed 12 month delay between publishing the first paper on a structure and making the coordinates public.

With experiments carried out over an extended period of time, (e.g. population based studies), it is reasonable to expect that subsets of data analysed by the investigator(s) be made available for sharing. The investigator(s) can then continue to benefit from further reasonable periods of exclusive analysis while the dataset as a whole matures.

Research involving human participants

Investigators carrying out research involving human participants must ensure that consent is obtained to share information; furthermore the necessary legal, ethical and regulatory permissions regarding data sharing should be in place prior to disclosing any data. Every effort must be made to protect the identity of participants and, prior to sharing, data should be anonymised. In addition, any indirect identifiers that may lead to deductive disclosures should be removed to reduce the risk of identification. In most instances, sharing data should be possible without compromising the confidentiality of participants but if there are circumstances where data needs to be restricted due to the inability to protect confidentiality this should be fully addressed in the data management and sharing plan.

Data sharing requests

When a principal investigator is contacted with a request to share his/her data, they may ask the requestor to provide a brief research proposal on how they wish to use the data. It could include the objectives, what data are requested, timelines for use, intellectual property and publication rights etc. This may form the basis of a data sharing agreement (see below). If the principal investigator has doubts over scientific validity of the proposal or the requestor's ability to analyse/interpret data correctly, this should discussed with the requestor. A refusal to share data in such circumstances must have clear justification.

Data sharing agreements

To ensure that data are used appropriately investigators may consider implementing a data sharing agreement that indicates the criteria for data access and conditions for research use. This can ensure the responsibilities of both parties, along with intellectual property, citation and publication rights are agreed at the outset. It may incorporate privacy and confidentiality standards, as needed, to ensure data security at the recipient site and prohibit manipulation of data. For further guidance on managing data access and the development of data sharing agreements please refer to the 'Samples and Data for Cancer Research: Template for Access Policy Development' document(link is external).

Data acknowledgement

As a minimum, researchers using shared data are expected to acknowledge the investigators who generated the data upon which any published findings are based. When both parties have collaborated using a shared dataset, coauthorship on publications may be more appropriate. Researchers using shared data are also expected to acknowledge Cancer Research UK for supporting the original study.

Data preservation

Once the funding for a project has ceased researchers should preserve all data resulting from that grant to ensure that data can be used for followup or new studies. We expect that data be preserved and available for sharing with the science community for a minimum period of five years following the end of a research grant.

Cancer Research UK Data sharing FAQ

What is CRUK's data sharing and preservation policy?

The policy states our expectation that the researchers we fund should make their research data as widely available and with as few restrictions as possible, while maximising patient benefit. It also highlights the need for all researchers to plan how they will manage and share their data. Investigators will be asked to provide a data management and sharing plan as part of a funding application.

To whom does this policy apply?

It is applicable to all candidates seeking funding from CRUK after 1 April 2009.

Who benefits from data sharing?

Managed data sharing can benefit investigators, the wider scientific community, funding agencies and the public. We believe that helping to make research data more readily available will reinforce open scientific enquiry, stimulate new investigations and analyses and thus maximise the value of the research we fund.

Which scientific areas and type(s) of data can be shared?

Data can be shared in all research areas where it is cost effective. There is a particularly strong scientific case with studies that generate large volumes of data that may yield further findings from analysis outside the scope of the original investigation.

Does data sharing apply only to published data?

No, the policy encompasses all high quality data from funded research that can be shared regardless of whether they have been used in a publication.

How does the data sharing policy relate to CRUK's policies on intellectual property?

We support the appropriate protection and use of patents and other intellectual property rights to maximise the opportunity to benefit patients. We expect our researchers (with the support of our Commercial Partnerships team) to manage and protect the intellectual property in their research, so that it can be used for public benefit. The data management and sharing policy does not alter this requirement. Data should not be shared or disclosed before a patent can be filed on an invention arising from research it funds. If any researcher suspects that an invention has or may be made they are expected to notify their technology transfer office and our Commercial Partnerships team, and to defer sharing any relevant data until the situation has been reviewed.

My research is translational and needs to be protected and/or commercialised. How do I deal with data sharing?

We recognise that certain types of research, particularly with a translational focus, are more likely to result in patentable inventions which can be developed further to deliver patient benefit, or to result in commercial collaborations. As such, you should keep our Commercial Partnerships team informed of all developments where there might be a potential for commercial interest, and in particular prior to

any data sharing. It may be necessary to delay data sharing and modify any data sharing plan to ensure that patient benefit can be maximised.

My research seeks supports from both the public and private sectors. How do I deal with the sharing of data?

Where research is funded by a commercial sponsor, restrictions on data sharing may apply in arrangements agreed with the sponsor. Any such restriction(s) should be highlighted in the data management and sharing plan. In the event that you apply for or receive commercial funding for any part of research that we support, you should advise our Commercial Partnerships team of the situation without delay.

What is the timescale for sharing data? Can I delay sharing until publication?

The latest point at which data should be shared is:

- Acceptance for publication of the results upon which the data is based, when no third party agreements restrict sharing.
- At a defined point that is the accepted procedure for the research area. For example with crystallography, there is an agreed 12-month delay between publishing the first paper on a structure and making the co-ordinates public.
- After all relevant patents are filed or a decision is made not to file a patent.

Investigators should aim to release data earlier than this if possible - particularly when it would be of benefit to the wider research community.

What do I need to include in my applications and where do I put the information about data sharing?

A data management and sharing plan should include concise plans for data sharing and the timeframe or explain why data sharing is not possible or appropriate. It is however recognised that plans for data sharing and timeframes might change during the course of the research and any such changes should be communicated to CRUK.

Further information can be found in our Data Sharing Guidelines. A box for completing a data management and sharing plan is incorporated into the grant application form or, for Population Research Committee schemes, provided as a separate document.

How will my data sharing plan be assessed and will it affect the outcome of my application?

It will be assessed by the funding committee. If the funding committee is not satisfied with the plan, you may be asked to make revisions before a grant award letter is issued.

Will CRUK provide funds for data sharing?

We regard the management and sharing of data generated through our funded research as a fundamental component of good scientific practice. Therefore, applicants may include proportionate, relevant data management and sharing activities as a running cost within applications. Funds will need to be fully justified and costed appropriately.

How will CRUK monitor adherence to the data sharing policy?

The funding committees monitor data management and sharing plans through the committee's grant review process and the end of grant report. However, we understand that an investigator may need to adapt the method and timelines for sharing during the course of the study - for example when intellectual property arises (see our Data sharing Guidelines).

How can I balance data sharing with the need to safeguard research participants?

All research involving human participants, or data or samples derived from human participants (such as cohort studies, clinical trials etc.), must include appropriate

safeguards to protect the privacy of research participants. You must ensure that the necessary patient consent is obtained prior to data sharing.
What are the responsibilities of researchers and others who access and use data? We believe that data sharing for the benefit of the research community as a whole will only proceed if those using the data also adopt good research practice. To ensure that data is used appropriately investigators may consider implementing a data sharing agreement that indicates the criteria for data access and conditions for research use. It may incorporate privacy and confidentiality standards, as needed, to ensure data security at the recipient site, protect intellectual property rights and prohibit manipulation of data.

Circle of Service Foundation

Payment Conditions.

(a) No Year One payment shall be due with respect to the Grant unless and until you submit an initial "Organizational Data-sharing Plan." The Organizational Data-sharing Plan should detail how you plan to encourage and/or require funded researchers to (1) craft data-sharing plans and (2) share applicable data that are generated from their funded projects.

(b) No Year Two payment shall be due with respect to the Grant unless and until you submit an update on the progress towards finalizing and implementing the Organizational Data-sharing Plan.

Fondation Leducq

Leducq networks can provide advantages to members in a number of ways, including:

o Bringing new approaches to old problems;
o Catalyzing creativity through multidisciplinary interaction; o Increasing efficiency through the participation of members with complementary specialized skills;
o Expanding access to resources, such as genetically modified animals, specialized techniques, supplies, equipment, etc.;
o Allowing for the rapid testing of hypotheses in different model systems;
o Enhancing information sharing and communication, particularly with regard to virtual and interactive working methods, and access to databases of mutual interest
o Promoting personnel exchange, especially for early-career investigators, over the short, medium and long term. Training positions may be made available to researchers from other network member institutions;
o Developing joint research infrastructures and adapting existing equipment for shared use;
o Optimizing the use of support staff and associated personnel among members of the network.
o Provide an enhanced training environment for early career investigators.

Foundation Fighting Blindness

On grant application info: require "Research Project Description including Data Sharing plan"

PLOS Journals

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication.

Methods acceptable to PLOS journals with respect to data sharing are listed below, accompanied by guidance for authors as to what must be indicated in their data availability statement and how to follow best practices in reporting. If authors did not collect data themselves but used another source, this source must be credited as appropriate. Authors who have questions or difficulties with the policy, or readers who have difficulty accessing data, are encouraged to contact the journal office (plosone@plos.org). If you have broader questions about the PLOS data availability policy, contact data@plos.org.

The data policy was implemented on March 3, 2014. Any paper submitted before that date will not have a data availability statement. However for all manuscripts submitted or published before this date, data must be available upon reasonable request.

 Download the full text of the older policy (PDF).

Acceptable Data-Sharing Methods

Data deposition (strongly recommended)

All data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository, unless already provided as part of the submitted article. Repositories may be either subject-specific (where these exist) and accept specific types of structured data, or generalist repositories that accept multiple data types, such as Dryad. Guidance on acceptable repositories is included below.

The Data Availability Statement must specify that data are deposited publicly and list the name(s) of repositories along with digital object identifiers or accession numbers for the relevant data sets. Read more about accession numbers.

Data in supporting information files

For smaller data sets and certain data types, authors may upload data as supporting information files accompanying the manuscript. (See also additional information regarding appropriate use of supporting information files.) Authors should take care to maximize the accessibility and reusability of the data by selecting a file format from which data can be efficiently extracted (for example, spreadsheets are preferable to PDF when providing tabulated data).

If data deposition or provision in supporting information is not ethical or legal (i.e., underlying data pose privacy or legal concerns e.g., where data might reveal the identity or location of participants), the following two methods may be acceptable alternatives, subject to case-by-case evaluation:

Data made available to all interested researchers upon request

The Data Availability Statement must specify "Data available on request" and identify the group to which requests should be submitted (e.g., a named data access committee or named ethics committee). The reasons for restrictions on public data deposition must also be specified. Note that it is not acceptable for the authors to be the sole named individuals responsible for ensuring data access.

Data available from a third party

We consider third-party data to be data not owned by the authors. Authors should share any data specific to their analysis that they can legally distribute. If an author does not own the data set, they must include in the Data Availability Statement all necessary contact information where an interested researcher would need to apply to gain access to the relevant data.

If permission was required to use a third-party data set (e.g., very large unpublished genome data or similar), authors must include the third-party source and verification of permission in the Data Availability Statement, as well as proper acknowledgment in the manuscript.

Please note that authors are responsible for ensuring that data will be available from the data owner post-publication, in the same manner as the authors obtained the data.

Unacceptable Data Access Restrictions

PLOS journals will not consider manuscripts for which the following factors influence ability to share data:

- Authors will not share data because of personal interests, such as patents or potential future publications.
- The conclusions depend solely on the analysis of proprietary data, whether these data are owned by the authors, by their funders or institutions, or by other parties. We consider proprietary data to be data owned by commercial interests, or copyrighted data that the data owners will not share, e.g., data from a pharmaceutical company that will share the data only with regulatory agencies for purposes of drug approval, but not with researchers. If proprietary data are used and cannot be accessed by others (in the same manner by which the authors obtained them), the manuscript must include an analysis of public data that validates the conclusions so that others can reproduce the analysis and build on the findings.

See acceptable data access restrictions here.

Explanatory Notes and Guidance

A compilation of frequently asked questions about the PLOS Data Policy is available and is updated periodically.

Definition of data that must be shared

PLOS defines the "minimal data set" to consist of the data set used to reach the conclusions drawn in the manuscript with related metadata and methods, and any additional data required to replicate the reported study findings in their entirety. Authors do not need to submit their entire data set if only a portion of the data were used in the reported study. Also, authors do not need to submit the raw data collected during an investigation if the standard in the field is to share data that have been processed.

Please note that PLOS does not permit references to "data not shown." Authors should provide the relevant data within the manuscript, the Supporting Information files, or in a public repository. If the data are not a core part of the research study being presented, we ask that authors remove any references to these data.

Guidance on sharing data sets that derive from clinical studies or other work involving human participants

For studies involving human participants, data must be handled so as to not compromise study participants' privacy. PLOS recommends that researchers follow established guidance and applicable local laws in ensuring they do not compromise participant privacy. Resources which researchers may consult for guidance include:

- US National Institutes of Health: Protecting the Rights and Privacy of Human Subjects
- Canadian Institutes of Health Research Best Practices for Protecting Privacy in Health Research
- UK Data Archive: Anonymisation Overview
- Australian National Data Service: Ethics, Consent and Data Sharing

Steps necessary to protect privacy may include de-identification, blocking portions of the database, or license agreements directed specifically at privacy concerns. Authors should indicate, as part of the ethics statement, the ways in which the study participants' privacy was preserved. If license agreements apply, authors should note the process necessary for other researchers to obtain a license.

Recommended Repositories

PLOS requires that authors comply with field-specific standards for preparation and recording of data and select repositories appropriate to their field, for example deposition of microarray data in ArrayExpress or GEO; deposition of gene sequences in GenBank, EMBL or DDBJ; and deposition of ecological data in Dryad. Authors are encouraged to select repositories that meet accepted criteria as trustworthy digital repositories.

PLOS has identified a set of established repositories below, which are recognized and trusted within their respective communities. For further information on environmental and biomedical science repositories and field standards, we suggest utilizing FAIRsharing; we have also created a FAIRsharing page of PLOS recommended data repositories. Additionally, the Registry of Research Data Repositories (Re3Data) is a full scale resource of registered repositories across subject areas. Both FAIRsharing and Re3Data provide information on an array of criteria to help researchers identify the repositories most suitable for their needs (licensing, certificates and standards, policy, etc.).

Authors are encouraged to select the repository most appropriate for their research. PLOS does not dictate repository selection for the data access policy. If authors use repositories with stated licensing policies, the policies should not be more restrictive than the Creative Commons Attribution (CC BY) license. More information about the content license can be found in our licenses and copyright policy.

If no specialized community-endorsed open repository exists, institutional repositories that use open licenses permitting free and unrestricted use or public domain, and that adhere to best practices pertaining to responsible data sharing, sustainable digital preservation, proper citation, and openness are also suitable for data deposition.

FAQs for Data Policy

Policy overview

Why does PLOS have a data policy?

PLOS believes that making data available fosters scientific progress. Data availability allows and facilitates:

- Validation, replication, reanalysis, new analysis, reinterpretation or inclusion into meta-analyses
- Reproducibility of research
- Efforts to ensure data are archived, increasing the value of the investment made in funding scientific research
- Reduction of the burden on authors in unearthing old data, retaining old hard drives and answering email requests
- Easier citation of data as well as research articles, enhancing visibility and ensuring recognition for authors

PLOS understands that some authors may not want to share data, just as some choose not to make their articles available Open Access, but we believe that authors publish their work precisely in order to allow others to benefit from it. More importantly, researchers want to see their work used and cited by others.

Exceptions

What are the acceptable exceptions to making the data publicly available?

We hope that data will be publicly available to all interested researchers, but we do understand that ethical and legal restrictions may prohibit this. The policy is not intended to overrule local regulations, legislation or ethical frameworks. Where these frameworks prevent or limit data release, authors should make these limitations clear in the Data Availability Statement at the time of submission.
Possible exceptions to making data publicly available include:

- Data cannot be made publicly available for ethical or legal reasons, e.g., public availability would compromise patient confidentiality or participant privacy.
- Data deposition could present some other threat, such as revealing the locations of fossil deposits, endangered species, or farms/other animal enclosures etc.

We hope that institutions will recognize the importance of preserving data and making it available, especially given concerns over data preservation and reproducibility, and that they will support their researchers in making data available. We encourage researchers and their institutions to consider whether a Data Access Committee could be convened to hold data and respond to requests for data. Since many institutions do not have committees in place to help with this process, we will work with authors to try to identify a solution in the meantime.
Please contact the journal office (plosone@plos.org) to discuss:

- if you feel unable to share data for reasons not specified above, or
- if you have concerns about the ethics or legality of sharing your data.

My study uses proprietary data; what should I do?

We consider proprietary data to be data owned by commercial interests, or copyrighted data that the data owners will not share, e.g., data from a pharmaceutical company that will share the data only with regulatory agencies for purposes of drug approval, but not with researchers.
PLOS will not consider submissions where the conclusions depend solely on the analysis of proprietary data, whether these data are owned by the authors, by their funders or institutions, or by other parties. If proprietary data are used and cannot be accessed by others (in the same manner by which the authors obtained it), the

manuscript must include an analysis of public data that validates the conclusions so that others can reproduce the analysis and build on the findings.

Clinical data

My study analyzes qualitative data and the participants did not consent to have their full transcripts made publicly available. What should I do?

The data policy exception related to privacy concerns pertains in this case. However, if requested, at least the excerpts of the transcripts relevant to the study would need to be shared. In this case, authors should include the contact information where requests may be sent in their Data Availability Statement, and state that excerpts of data are available on request. If even sharing excerpts would violate the agreement to which the participants consented, then please inform the journal office.

My study was conducted in humans and my minimum data set includes information on individuals. What should I do?

Adherence to the PLOS data policy must never breach patient confidentiality. Authors should ensure that the data shared are in accordance with patient consent. Authors should provide only the data that are used in the specific study. Individual patient data should not include the following personal data (see *Publication and access to clinical-trial data* and *Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers*):

- name, initials, address including full or partial postal code;
- telephone or fax numbers of contact information, email address;
- unique identifying numbers, vehicle identifiers, medical device identifiers;
- Web or internet protocol addresses;
- biometric data, facial photograph or comparable image, audiotapes;
- names of relatives;
- dates related to an individual, including birthdate.

The following may not be appropriate to include depending on what other information is provided:

- place of treatment or health professional responsible for care;
- gender;
- rare disease or treatment;
- sensitive data, such as illicit drug use or risky behavior;
- place of birth;
- socioeconomic data, such as occupation or place of work, income, or education household;
- family composition;
- anthropometric measures;
- number of pregnancies;
- ethnicity;
- year of birth or age;
- verbatim transcripts or responses.

Also potentially inappropriate to include, depending on the type of information provided, are data on population sizes of less than 100 or those with small numerators, e.g., event counts less than 3. (Information from Hrynaszkiewicz I, Norton M L, Vickers A J, Altman D G. (2010). Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. BMJ 340:c181. http://www.bmj.com/content/340/bmj.c181.long)

The data from my study relates to a potential medicine that will be submitted to the European Medicines Agency (EMA) for approval. Do I need to wait until after the approvals process to make the data from my article available?

The data shared according to the PLOS policy likely represents only a small proportion of the evidence submitted to the EMA for approval and so should not interfere with approvals processes of the EMA. The EMA's Policy 0070, that specifies data release after authorization, applies only to the data held by the regulatory agency submitted as part of a marketing authorization application. As such, the PLOS data policy is fully compatible with the data sharing polices of the EMA. Therefore, authors should make the data underlying the findings described in their manuscript available at the time of publication.

Genetic data

See additional FAQs about submitting genetic data.

See the related *PLOS Genetics* editorial about genetics submissions to PLOS journals.

Depositing data

What if I cannot provide accession numbers or DOIs for my data set at submission?

If this is the case, authors may submit their manuscript and note in their Data Availability Statement that their accession numbers or DOIs will not be made available until acceptance. The journal office will contact you at acceptance to provide this information, and will hold your paper upon acceptance until we receive these identifiers for your data set.

Providing 'private' access for reviewers and editors during the peer review process is acceptable. Many repositories permit private access for review purposes, and have policies for public release at publication. If this is not possible, authors can provide the data via other means, such as zipped files via email, Dropbox etc. Please contact the journal office (plosone@plos.org) for assistance.

Is PLOS integrated with any repositories?

PLOS has a Data Repository Integration Partner Program that integrates our submission process with partner data repositories to better support data sharing and author compliance of the PLOS data policy. Our submission system is integrated with partner repositories to ensure that the article and its underlying data are paired, published together and linked. The integration facilitates deposition of data alongside article submission, which may also facilitate consideration and peer review of submissions.

Current partners include Dryad and FlowRepository. We are expanding the current selection of partners to integrate with more data centers. PLOS is repository agnostic; provided that data centers meet our baseline criteria (license and availability, reliability, preservation) that ensure trustworthiness and good stewardship of data, we would accept data submitted in those locations.

Partner repositories may have a data submission fee. PLOS is not able to cover this fee and authors are under no obligation to use any specific repository. PLOS does not gain financially from our association with any integrated partners. More information on the program can be found here.

How do I deposit data with a data repository integration partner?

Once an author deposits data in the integrated repository, s/he receives a provisional data set DOI along with a private reviewer URL link. Upon submission to PLOS, authors must include the data DOI into the Data Availability Statement. They should also provide the reviewer URL, which will permit restricted access to the data during peer review. If a manuscript is editorially accepted by a PLOS journal, the publication of the article and public release of the data set will be automatically coordinated.

I cannot afford the cost of depositing a very large amount of data. What should I do?

PLOS encourages authors to investigate all options and to contact their institutions if they have difficulty providing access to the data underlying the research. Authors facing these challenges are encouraged to submit their manuscript and PLOS will work with them to find a solution. If this is the case, please email the journal office.

What are acceptable licenses for my data deposition?

Data should be covered by a CC BY license or a less restrictive license.

Submitting to PLOS

What is the data availability statement and what should I write?

Upon submission to a PLOS journal, authors are asked to enter the location and availability of their data in the submission system. What is written in this text box will be published as is, should the paper be accepted.

If data are freely available, we ask that authors note this and state the location of their data:

- Within the paper, supporting information files, in a public repository (include DOI, accession)

If data are freely available and owned by a third party, please state:

- The owner of the data set where requests may be sent to

Note: If data have been obtained from a third party, we require that any researcher will be able to obtain the data set in the same manner by which the authors obtained it.

If there are any approved restrictions on the data set, for ethical or legal reasons, please state:

- The availability of the data;
- A brief description of the ethical or legal restrictions on the data set;
- A contact to whom requests for the data may be sent.

What data are required and what is meant by minimal data set?

PLOS defines the "minimal data set" to consist of the data set used to reach the conclusions drawn in the manuscript with related metadata and methods, and any additional data required to replicate the reported study findings in their entirety. Authors do not need to submit their entire data set, or the raw data collected during an investigation. Please submit the following data:

- The values behind the means, standard deviations and other measures reported;
- The values used to build graphs;
- The points extracted from images for analysis.

NOTE:

- Authors are not required to make all images available, but we do require a sample Western Blot, Immunohistochemistry image, fMRI image, etc. to be included with the submission files or in a public repository.
- Please note that PLOS does not permit references to "data not shown." Authors should provide the relevant data within the manuscript, the Supporting Information files, or in a public repository. If the data are not a core part of the research study being presented, we ask that authors remove any references to these data.

What format should I use for my data?

The file format used to submit data should follow the standards in the field. If there are currently no standards in the field, please submit the data in an accessible format from which data can be efficiently extracted (e.g., Excel rather than PDF).

How do I submit data as supporting information files?
Upon submission and at revision, authors have the opportunity to upload supporting information files. There is a 10 MB limit per file, but that is unlikely to be exceeded with Excel files or anything similar. If the files do exceed this amount, authors should zip or otherwise compress the files before submission.
In choosing between supporting information files and a repository, please refer to our blog post on uses of supporting information files.
What if data are found to not be accessible or other issues are found after publication?
PLOS will follow up with the authors and take action as necessary. PLOS reserves the right to issue corrections, notifications or retractions when authors do not comply with our policies.

Nature Journals

Availability of Data

Supporting data must be made available to editors and peer reviewers at the time of submission for the purposes of evaluating the manuscript. All manuscripts reporting original research published in Nature Research journals must include a data availability statement (see http://www.nature.com/news/announcement-where-are-the-data-1.20541). The statement should be placed at the end of the Methods section; for papers that do not have a Methods section, data availability statements should be provided as a separate section before the References or Acknowledgements, whichever comes first. This policy is effective on papers (new submissions and revisions) submitted to all Nature Research journals. For further guidance, please refer to the data availability and data citations policy information and Frequently Asked Questions (FAQs).
The preferred way to share large data sets is via public repositories. Details about how to share some specific data sets can be found in the sections below.
Some of these repositories offer authors the option to host data associated with a manuscript confidentially, and provide anonymous access to peer-reviewers before public release. These repositories then coordinate public release of the data with the journal's publication date. This option should be used when possible but it remains the author's responsibility to communicate with the repository to ensure that public release is made on time for online publication of the paper. For information about suitable public repositories, see sections that follow.
Unstructured repositories like figshare and Dryad are suitable alternatives if no structured public repositories exist. As a less desirable alternative, data sets can be made available as Supplementary Information files, which will be freely accessible on nature.com upon publication. In rare cases when data files cannot be deposited in an accessible repository for technical reasons, authors must make the data available to editors and peer reviewers if requested. After publication, authors must likewise arrange to make the data available to any reader directly upon reasonable request.
Nature Research journals encourage authors to consider the publication of a Data Descriptor in *Scientific Data* to increase transparency and enhance the re-use value of data sets used in their papers. Data Descriptors are designed to be complementary to a primary paper and can be published prior to, simultaneously,

or after publication of the primary paper. Nature Research journals will not consider prior Data Descriptor publications to compromise the novelty of new manuscript submissions, as long as those manuscripts go substantially beyond a descriptive analysis of the data and report important new scientific findings appropriate for the journal. (This policy does not necessarily extend to journal articles whose primary purpose is to describe a new data set or resource.)
Nature Research journals' data availability policies are compatible with the standardised research data policies set out by Springer Nature.
Further reading

Mandates for specific datasets

For the following types of data set, submission to a community-endorsed, public repository is mandatory. Accession numbers must be provided in the paper. Examples of appropriate public repositories are listed below.

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)

Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Microarray data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database
Proteomics data	PRIDE

Special considerations

DNA and protein sequences: When publishing reference genomes, the assembly must be made available in addition to the sequence reads. Sequence must be deposited even for short stretches of novel sequence information such as epitopes, functional domains, genetic markers, or haplotypes. Short novel sequences must include surrounding sequence information to provide context. The sequences of all small RNA probes central to the conclusions of the paper must be provided.

Linked phenotype and genotype data for human subjects: should be submitted to a public repository with appropriate access control (see above). Any restrictions on data access for sensitive data (for example electronic medical records, forensic data, and personal data from vulnerable populations) require an explanation of the nature of and reasons for the restrictions, and details of the conditions under which the data can be accessed or reused. (See the related *Nature Genetics* Editorial discussing privacy issues.)

Macromolecular structures: Official validation reports from the wwPDB are required for peer review. Atomic coordinates and related experimental data

(structure factor amplitudes/intensities for crystal structures, or restraints for NMR structures) must be provided upon request. Electron microscopy-derived density maps and coordinate data must be deposited in EMDB. Accessibility in repositories must be designated "for immediate release on publication."
Crystallographic data for small molecules: Manuscript reporting new three-dimensional structures of small molecules from crystallographic analysis should include a .cif file and a structural figure with probability ellipsoids for publication as Supplementary Information. The structure factors for each structure should also be submitted. Both the structure factors and the structural output must have been checked using the IUCR CheckCIF routine, and a PDF copy of the output must be included at submission, together with a justification for any alerts reported.
Recommendations for other datasets
In addition to these mandates, the preferred way to share any data sets is via public repositories. *Scientific Data*, a sister publication to Nature Research journals, maintains a list of approved and recommended data repositories organized by discipline. Please consult this list to identify an appropriate repository for your data sets.
When repositories do not exist for a particular data type, authors can deposit and share data via figshare or Dryad, two general-purpose scientific data repositories.

Wiley Publishing

Sharing and Citing your Research Data
In the academic community there is an increased pressure on researchers to share and archive their data, with many funders now mandating data publication. The sharing of data enables others to reuse experimental results and supports the creation of new work built on previous findings, improving the efficiencies of the research process and supporting the critical goals of transparency and reproducibility.

At Wiley, we support the growing movement to make science more open, because this leads to a fairer, more efficient and accountable research landscape, which will ultimately drive a more effective and faster pace of discovery. We are committed to improving openness, transparency, and reproducibility of research. Fundamental to enabling reproducible research is the easy access to and ready discovery of its supporting data, made possible through a robust and universal framework that allows research data to be cited through standard reference lists. This will ensure that data is treated as a first-class research object, easily accessible as part of the scholarly literature, and that researchers are credited for their work.

Click image to enlarge

Wiley is actively involved in contributing the research data community as an organizational member or signatory to the following initiatives:

- **Research Data Alliance (RDA) Organizational Member**
- **International Council for Science World Data System (ISCU-WDS) Associate Member**
- **ORCID Organizational Member**
- **Initiative for Open Citations (I4OC) Participating Publisher**
- **Transparency and Openness Promotion (TOP) Guidelines Organizational Signatory**
- **STM Brussels Declaration Organizational Signatory**
- **FORCE 11FAIR Data Principles, Endorsed**
- **Joint Declaration of Data Citation Principles (JDDCP), Organizational Signatory**

Wiley's Data Sharing Policies
Authors of articles published in Wiley journals are encouraged to share their research data including, but not limited to: raw data, processed data, software, algorithms, protocols, methods, materials.

The majority of Wiley's journals enforce one of the following standardized data sharing policies:

Encourages Data Sharing
"[Journal] encourages authors to share the data and other artefacts supporting the results in the paper by archiving it in an appropriate public repository. Authors should include a data accessibility statement, including a link to the repository they have used, in order that this statement can be published alongside their paper."

Expects Data Sharing
"[Journal] expects that data supporting the results in the paper will be archived in an appropriate public repository. Whenever possible the scripts and other

artefacts used to generate the analyses presented in the paper should also be publicly archived. Exceptions may be granted at the discretion of the editor for sensitive information such as human subject data or the location of endangered species. Authors are expected to provide a data accessibility statement, including a link to the repository they have used, to accompany their paper."

Mandates Data Sharing
"[Journal] requires, as a condition for publication, that the data supporting the results in the paper will be archived in an appropriate public repository. Whenever possible the scripts and other artefacts used to generate the analyses presented in the paper should also be publicly archived. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species. Authors will be required to provide a data accessibility statement, including a link to the repository they have used, for all accepted papers."

See below for Wiley's recommended methods of choosing an appropriate data repository for your research:

- **Visit our Author Compliance Tool to check the data sharing policy of your chosen journal and/or funder before submitting your work**
- **Visit re3data.org or fairsharing.org to help identify registered and certified data repositories relevant to your subject area**

Wiley's Data Citation Policy
In recognition of the significance of data as an output of research effort, Wiley has endorsed the FORCE11 Data Citation Principles.

Wiley journals require data to be cited in the same way as article, book, and web citations and authors are required to include data citations as part of their reference list.

Data citation is appropriate for data held within institutional, subject focused, or more general data repositories. It is not intended to take the place of community standards such as in-line citation of GenBank accession codes.

When citing or making claims based on data, authors must refer to the data at the relevant place in the manuscript text and in addition provide a formal citation in the reference list. We recommend the format proposed by the Joint Declaration of Data Citation Principles:
[dataset] Authors; Year; Dataset title; Data repository or archive; Version (if any); Persistent identifier (e.g.DOI)

Add [dataset] immediately before the reference so we can properly identify it as a data reference. The [dataset] identifier will not appear in your published article.

Data Sharing and Citation Policies FAQs
1. **What research outputs are classified as data?**
2. **Why should I share my research data?**
3. **Where can I archive my data?**

4. What is a data accessibility statement?
5. Why should I cite research data?
6. How do I cite research data?
7. What journals do these policies apply to?
8. Is it mandatory to share my data for every article?
9. Do the policies apply to sensitive or confidential data and/or data subject to third party restrictions?
10. What are the copyright/license implications for sharing data?
11. When should research data be shared?
12. Will research data publically posted ahead of submission be considered prior publication?
13. Are my data files subject to peer/editorial review?
14. Can my manuscript be rejected on the basis of my data files?
15. Will my manuscript be rejected if I do not submit data files?

What research outputs are classified as data?

Definitions of what research data is varies by discipline. 'Data' includes a research output that has been collected, observed or created for the purpose of analysis to produce the research results. Research data can include (but are not limited to): raw data, processed data, software, algorithms, protocols, methods, materials, photographs, specimens, etc. Generally, these policies apply to all research data that underlie and support the results documented in research articles. However, journals or communities might have more specific standards.

Why should I share my research data?

1. **Funders increasingly ask researchers to make their data publically available. According to SHERPA/JULIET, which tracks OA funders' policies on data sharing, there are a growing number of funders who encourage it.**
2. **Further, opening access to the world's research data offers huge potential to improve the transparency of research, accelerate the pace of discovery, improve return on investment, and lead to a future in which more research can be independently verified or made reproducible.**
3. **Wiley is committed to building and supporting connections between researchers and research communities, to improving the discoverability and reproducibility of research, and to encouraging openness and transparency in the exchange of knowledge and information.**

Where can I archive my data?

Choosing where to publish your datasets can be problematic and time consuming. See below for Wiley's recommended methods of choosing an appropriate data repository for your research:

- **Visit re3data.org or for extensive catalogues of registered and certified data repositories**

- **Some funders have designated archives set up for researchers to deposit their data. Check our Author Compliance Tool to see your funder's data sharing policy.**

In general, research data should be submitted to discipline-specific, community-recognized repositories where possible, or to general-purpose repositories if no suitable community resource is available. If your funder or target journal do not have specific data repository recommendations, researchers from all disciplines can consider generalist repositories such as Dryad, figShare, or Zenodo.

What is a data accessibility statement?
Data accessibility statements provide information about where the research data and other artefacts supporting the results reported in the paper can be found. Where applicable, links to the repository where the dataset(s) are publicly archived are included. Wiley's data sharing policies either recommend or require (depending upon policy) the inclusion of a data accessibility statement as part of the manuscript. Some funders require data accessibility statements be included in publications, authors must confirm any funder-specific requirements.

Why should I cite research data?
Wiley is implementing the FORCE 11 Joint Declaration of Data Citation Principles – this means that authors are required to include data citations as part of their reference list and Wiley journals require data to be cited in the same way as article, book, and web citations.

Assigning a persistent identifier to your research data enables other researchers to cite your data, as well your published research article. Formal citation in reference lists supports reproducibility, facilitates the tracking of data reuse, and may help recognize or credit individual's contributions to research and the work put into collecting, managing, and archiving data.

How do I cite research data?
We recommend the format proposed by the Joint Declaration of Data Citation Principles:

[dataset] Authors; Year; Dataset title; Data repository or archive; Version (if any); Persistent identifier (e.g. DOI)

We have altered our production and publication systems to process data citations. By adding [dataset] before the reference, our systems will recognize the citation appropriately. This additional tag will not be visible within the reference list of the published article. Readers will therefore enjoy the same benefits as for article citations, including the ability to easily navigate to where the work was cited in the article and quickly access the referenced material via direct links.

What journals do these policies apply to?
Wiley is encouraging all journals to adopt one of the standard research data sharing policies. All Wiley journals are implementing the data citation policy.

You can access the list of journals and the policies they support at the Author Compliance Tool.

Is it mandatory to share my data for every article?
The minimum requirement of all policy types is to encourage data sharing; only those journals that have adopted the strongest level of data sharing policy mandate data sharing for every article. And, all policy types recognize that some data (such as data about identifiable human research participants), can't be openly shared. You can access the list of journals and the policies they support at the Author Compliance Tool.

Do the policies apply to sensitive or confidential data and/or data subject to third party restrictions?
Exceptions to policy and restrictions on data availability are granted for reasons associated with the protection of human privacy, issues such as biosafety, and/or to respect terms of use for data obtained under license from third parties. Confidential data, e.g., human subject or patient data, should always be anonymized, or permission to share should be obtained in advance. If in doubt, authors should seek counsel from their institution's ethics committee.

What are the copyright/license implications for sharing data?
Where data are held in repositories, the choice of license will be determined by the terms of the repository. Some funders also have specific license requirements. Authors are responsible for reviewing the license agreements during submission.

Researchers should ideally decide how their research data is made available, but can only share data they are legally permitted to share or make public. In general, a license that enables the maximum potential for reuse, such as one of the Creative Commons licenses (CC-0, -BY, -BY-NC), is preferred. It is the responsibility of the author depositing data to confirm they have the necessary rights to submit data to a repository or journal.

When should research data be shared?
Authors are encouraged to make research data available as early as possible, in accordance with community practice and as required by funder and institutional policy. Practice varies by field, and embargoes on data sharing are common practice in some communities so, in the absence of funder mandate, the relevant community standards should prevail. Only the "mandates data sharing" policy requires data sharing as a condition for publication and requires data sharing upon acceptance by the journal – authors should confirm the policy of their target journal prior to submission.

Will research data publically posted ahead of submission be considered prior publication?
Wiley does not generally consider research data deposit as prior publication, however individual journal policies may vary and we recommend that researchers contact their chosen journal's Editor if they are in doubt.

Are my data files subject to peer/editorial review?
If your data is available during peer review, it may be accessed by reviewers to help in their evaluation. Journal Editors likewise may use available data just as they would any other available resources.

Can my manuscript be rejected on the basis of my data files?
Conceivably, yes, if the reviewers and Editor(s) feel there are discrepancies between the data files (if checked) and the figures, tables, and graphs in your article.

Will my manuscript be rejected if I do not submit data files?
An Editor may choose to reject your manuscript if you are unwilling (rather than unable) to comply with the data sharing policy of the journal in question. View your chosen journal's data policy information on our Author Compliance Tool.

Wiley's Data Sharing Service
A number of Wiley journals participate in Wiley's Data Sharing Service, which enables you to automatically archive your data when submitting your article within the existing manuscript submission workflow. Wiley's Data Sharing Service is currently available through a partnership with figshare. Once accepted for publication, data files will be transferred automatically and deposited to the figshare data repository, without charge or further work. For more information, please visit our Data Sharing Service FAQs page.