



UK Dementia
Research Institute

UK DRI Data Sharing Policy

Founding funders:



Medical
Research
Council



Alzheimer's
Society



Alzheimer's
Research
UK

Table of Contents

1. Purpose	2
2. Background	2
3. Policy Statements	2
4. Guidelines	4
4.01. Data Covered by this Policy	4
4.02. Research with Human Participants	4
4.03. Exceptions to Data Release	4
4.04. Timeliness for Data Sharing	4
4.05. Standards, Metadata and Documentation	5
4.06. Method for Data Sharing and Open Data Licensing	5
4.07. Secondary Use and Rights of Data Provider	6
4.08. Data Sharing Support and Reporting Mechanism	6
Reference	7
Data Repositories	8

1. Purpose

The UK Dementia Research Institute (UK DRI) is a data-driven institute. We expect our researchers to adopt good research practice and act with integrity and transparency in collecting, managing, sharing and (re)using high quality, scientifically robust data. The UK DRI is committed to open science and strives to:

- Nurture a research environment that makes data open in a timely and responsible manner.
- Recognise the importance of data citation and credit acknowledgement.
- Use transparent, robust and fair processes to make decision concerning data openness.

This document covers the general principles, policy statements and guidelines on the public sharing of UK DRI data, including computer code, with the scientific community and is maintained by the UK DRI Informatics Steering Committee with the approval of the Scientific Strategy and Operational Board. **The policy applies to all UK DRI members¹ regardless of the funding source.** Where a research project includes external collaborator(s), it is important to have the [UK DRI headquarters \(HQ\) team](#) review the project agreement to ascertain any conflicting provisions. Collaborator(s) should be made aware of the UK DRI Data Sharing Policy and comply where applicable. In all instances data should be shared in a timely and responsible manner, preferably in line with this policy.

2. Background

The UK DRI is dedicated to advancing dementia research that will transform lives. Our researchers are generating large, diverse datasets that are important to both the primary researchers and the wider scientific community that may wish to reproduce and/or drive investigations in new directions. We recognise the importance of contributing to the growing global efforts in open science and believe data sharing will reinforce research transparency and stimulate innovative analyses. We are committed to ensuring UK DRI data can be put to maximum use by the scientific community for health and societal benefit, where legally, commercially and ethically appropriate.

The basis of effective data sharing and reuse is outlined in the [FAIR Guiding Principles for scientific data management and stewardship](#) that emphasize machine-actionability and stipulate the Findability, Accessibility, Interoperability and Reuse of data. Briefly, data (or any digital object) and metadata (information about the object) should be easily *findable* by both humans and computers (enabling automatic discovery). Data should be *accessible* (via authentication and authorisation if required), be integratable with other data, and *interoperate* with applications or workflows for storage, processing and analysis. (Meta)data should be well described to allow replication and/or combination in different settings to maximise *reuse*.

Open science is only sustainable if scientific credit is given to all involved. We are committed to develop a research environment that enables and rewards data sharing. Fundamentally, data sharing should be led by the scientific community and driven by scientific need without stifling creativity. We expect all UK DRI members to practice and promote data sharing to create an open research culture.

3. Policy Statements

The need to safeguard intellectual property (IP) and commercialisation opportunities should not unduly delay or prevent data sharing. We expect data to be shared with as few restrictions/barriers as possible in a timely and responsible manner. The policy reflects the broad statements laid out below and is supported by [Guidelines](#) to assist with compliance. It is consistent with our position on open science and our policies on [IP and Commercialisation](#) as well as [Collaboration](#).

- All UK DRI members should comply with the terms of the UK DRI Data Sharing Policy.
- All final research data (for definition see [4.01](#)) and computer code (henceforth referred jointly as 'data') generated by UK DRI members are covered by this policy and should be shared [according to the FAIR Principles](#), excluding legal, commercial or ethical exceptions as agreed with the UK DRI HQ team. Researchers should also adhere to all relevant regulatory requirements, including those relating to information governance and the ethical use of data. [See [4.01](#) to [4.03](#)]
- Final research data would be released no later than the publication of the main findings, the submission of the main findings to a preprint repository, or 24 months after generation, whichever is

¹ Please refer to the [UK DRI Membership Policy](#) for definition or [contact the HQ team](#) for questions.



the earliest. Computer code should be released upon submission of the manuscript to a preprint repository or upon publication, whichever is the earliest. [4.04]

- Data should be suitably annotated and documented with metadata to avoid misinterpretation. Established standards for (meta)data collection, management, ontology and formats should be adopted wherever possible. Appropriate Data Working Groups will be set up to identify data type-specific (meta)data standards and repositories. [4.05]
- Data should be made available under an appropriate license that allows reuse with minimal restrictions and deposited, wherever possible, to established public repositories listed in [Registry of Research Data Repositories \(re3data\)](#). [4.06]
- Researchers who generate, preserve and share data should receive full recognition by secondary users, their academic institutions and funders for promoting open science. Users are expected to duly cite the source, preserve data confidentiality and observe the ethical and legal obligations pertaining to the data. [4.07]
- The UK DRI aims to support our researchers in the application of the UK DRI Data Sharing Policy. [4.08]
- We expect UK DRI researchers to incorporate data management plans as an explicit part of their research projects. All UK DRI-supported projects would be required to agree to the UK DRI Data Sharing Policy.
- After a project has ended, data should be retained for a specific time period according to each host university's data management policy.

4. Guidelines

These guidelines are intended to provide UK DRI researchers with practical guidance for implementing the UK DRI Data Sharing Policy. The guidelines will be periodically reviewed by the UK DRI Informatics Steering Committee. Please [contact the UK DRI HQ team](#) for questions or concerns on the policy or guidelines.

4.01. Data Covered by this Policy

Here we term final research data as recorded factual material necessary to document, support and validate research findings. For most studies, final research data will be a computerized dataset. The dataset might include both raw data and derived variables, which should be appropriately described in the documentation accompanying the dataset. Final research data does not include laboratory notebooks, partial datasets, preliminary analyses, publication drafts, future research plan or peer review reports.

The UK DRI Data Sharing Policy covers both **final research data** and **computer code** (henceforth jointly referred as 'data') generated by all UK DRI members regardless of the funding source.

4.02. Research with Human Participants

Personal data is data containing identifying information. Collection, use, storage and sharing of personal data of living human participants is covered by the Data Protection Act 2018, UK's implementation of the General Data Protection Regulation (GDPR), in addition to any regulatory requirement (on, for example, information governance and ethics) of the specific study. It is important to be aware that Data Protection Act also covers recordings and images of living patients and research participants where these contain, or are held with, identifying information. Data that is anonymised or pseudonymised better protects participants' privacy and is not subject to the same restrictions as personal data.

For research with human participants, researchers must have participants' consent to collect, use, store and share their personal data. Consent procedures should include provision for data sharing in a way that maximises the value of the data for wider research use, while providing adequate safeguards for participants. For more information on consent for data sharing, please see the [UK Data Service](#). Every effort must be made to protect the identity of participants, including through appropriate anonymisation procedures and managed access processes. Prior to sharing, data should be anonymised, and any indirect identifiers that may lead to deductive disclosures should be removed to reduce the risk of identification. **All appropriate ethical, legal and institutional regulatory permissions must be in place before the data can be shared.**

4.03. Exceptions to Data Release

While [FAIR-compliant](#) data should be the presumed default, we recognise that there may be commercial, legal or ethical constraints on the release of some data. Delays and exemptions **may** be granted for reasons such as:

- IP rights: Data release may be delayed for assessment and execution of an IP protection strategy; for example, if the data forms part of a patent application.
- Proprietary data: Data generated from projects co-funded by third parties such as pharmaceutical or biotechnology industries may have constraints on public disclosure.
- Studies with clear risk of participant reidentification, bioterrorism.

Where researchers wish to delay or exempt the data from release, they should first consult their group leaders, who, after deeming the request essential, should return the completed [UK DRI Data Release Delay/Exemption Request Form](#) to [the HQ team](#) for approval by the HQ team and, if necessary, Chair of the relevant Data Working Group (or the Informatic Steering Committee, where appropriate).

4.04. Timeliness for Data Sharing

To ensure researchers who generate the data ('data generators') have the first opportunity to fully exploit the data and publish their findings, data generators have reasonable right of privileged use for a limited, defined period.

All data however should be made publicly available [according to the FAIR Principles](#) following the timeline below:

- Final research data – no later than 24 months after generation, the submission of the main findings to a preprint repository, or the publication of the main findings, whichever is the earliest.
- Computer code – upon submission of the manuscript to a preprint repository or upon publication, whichever is the earliest.

In addition, delays in sharing a novel methodology to collect data will restrict others from applying the methodology (and reusing the resulting datasets) and delay the recognition due to the developers. We thus encourage our researchers to share novel methodologies and associated data at the earliest opportunity.

4.05. Standards, Metadata and Documentation

Standards for ontologies (eg consistent nomenclature for disease name and brain area)², (meta)data formats, supporting metadata/documentation, and data collection/management methodologies are fundamental to effective, robust data sharing. Data released for sharing should:

- Meet well established, domain relevant community standards.
- Be validated and verified in line with established best practice and be of high quality.
- Be accompanied by all necessary supporting metadata and/or documentation.
- Be sufficient to reproduce all results and figures generated using the dataset.

To maximise integrability, interoperability and reuse, data should be collected and managed using widely accepted methodologies and formats where available. For data to be readily findable and effectively reused, appropriately structured metadata and/or documentation should be recorded, using a recognised or de facto standard schema. Metadata should include information about (1) regulatory and ethical requirements on access and use; (2) version-controlled protocols for equipment use, [data provenance](#) and context of data collection. Documentation may include code comments, technical descriptions, project-build guidelines, audit trail supporting technical decisions, and software dependencies.

Identification of standards for different data types (eg transcriptomics, proteomics, software) will be led by the UK DRI Informatics Steering Committee and associated Data Working Groups. A list of recommendation will be hosted on [UK DRI Portal](#) and regularly reviewed to reflect new entries.

4.06. Method for Data Sharing and Open Data Licensing

Wherever possible, data should be deposited in an established, domain-appropriate public repository listed in [re3data](#), complying to all (meta)data standards expected in the chosen repository. Please see [Data Repositories](#) for a non-exhaustive list of general and domain-specific repositories.³ Data should be citable and locatable unambiguously through the use of [persistent unique identifiers](#) such as [Digital Object Identifiers \(DOIs\)](#).

Data can be shared via either open or managed access:

- Open access data – data that can be released into the public domain without restriction.
- Managed access data – data that can only be released under certain conditions with restrictions on use and re-distribution usually related to the terms of consent given by research participants. This type of data should be deposited in a controlled, secure repository with strict protocols governing how information is managed, stored and distributed (eg [European Genome-phenome Archive](#)).

We strongly encourage the use of open access repositories, and one should only opt for managed access if it is explicitly required for the datasets in question.

Data should always be made available under licence to inform secondary users who owns the data and on what terms they can be used. Researchers are strongly encouraged to opt for the most open licence to allow the widest possible scope for reuse and redistribution. [Creative Commons Attribution](#) (CC BY) license

² See, for example, [the EMBL-EBI Ontology Lookup Service](#).

³ This list is subject to regular review by the UK DRI Informatics Steering Committee and Data Working Groups. Any changes to the list will be announced on Slack and an up-to-date list will be maintained on [UK DRI Portal](#).

is widely used for datasets and open access publications (see [Wellcome](#) for example). For more information, [Open Definition](#) lists conformant open licenses for data and content; [Open Source Initiative](#) lists open-source licenses for software. Different repositories may use or prefer different licenses; [re3data](#) indicates the license used by each listed repository.

4.07. Secondary Use and Rights of Data Provider

To fully credit the intellectual contributions of researchers who generate, preserve and share data, secondary users must appropriately cite the data upon which any published findings are based in the method and/or reference section. A data citation should, wherever possible, include sufficient information to allow the exact version of the data to be located by using a persistent unique identifier such as DOIs. Where data are shared directly from the data generator, a co-authorship may be appropriate based on mutual agreement.

A data citation does not suggest the data generator endorses the secondary user or their use of the data. Users should abide by the terms and conditions under which the data is accessed and understand that the data may be subject to change, errors and sampling bias. Users should carefully read the metadata and documentation to fully understand the scope and original intent of the data to reduce the risk of misinterpretation.

4.08. Data Sharing Support and Reporting Mechanism

The UK DRI recognises that data sharing has time and cost implications such as the hiring of specialist staff and periodic reviews of progress towards FAIR-compliant data. The cost must be balanced with the benefits to the UK DRI's research portfolio as a whole. The UK DRI aims to enable data sharing by providing:

- Staff support via the HQ team and/or the UK DRI Data Science and Informatics Team and/or through partnership(s).
- Information and training on relevant common standards, guidelines, databases and resources.
- Development or identification of best-practice (meta)data/documentation standards through the UK DRI Informatics Steering Committee and associated Data Working Groups.

Generation of FAIR-compliant data should be acknowledged as a legitimate output of the research process and recognised as such in a researcher's professional profile. UK DRI Group Leaders and Pilot Award holders are requested to capture their data-sharing activities (including information such as persistent unique identifiers and where the data is shared) on researchfish® and UK DRI Annual Planning Exercise, if applicable.

Reference

The UK DRI Data Sharing Policy borrowed heavily from:

- Wellcome Sanger Institute: Open access science
<https://www.sanger.ac.uk/about/who-we-are/research-policies/open-access-science/>
- National Institutes of Health: NIH data sharing policy and implementation guidance
https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm
- Cancer Research UK: Data sharing guidelines
<https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines>
- Howard Hughes Medical Institute: Sharing published materials/responsibilities of HHMI authors
<https://hhmicdn.blob.core.windows.net/policies/Sharing-Published-Materials-Responsibilities-of-HHMI-Authors>
- National Institute for Health Research: NIHR position on the sharing of research data
<https://www.nihr.ac.uk/documents/nihr-position-on-the-sharing-of-research-data/12253>
- Biotechnology and Biological Sciences Research Council: Data sharing and data resources
<https://bbsrc.ukri.org/about/policies-standards/data-sharing-policy/>
<https://bbsrc.ukri.org/research/resources/#datasharing>
- UK Research and Innovation: Open research
<https://www.ukri.org/funding/information-for-award-holders/data-policy/>
- Royal Society: Data sharing and mining
<https://royalsociety.org/journals/ethics-policies/data-sharing-mining/>
- University of Reading: Preserving and sharing data
<https://www.reading.ac.uk/RES/rdm/preservingandsharing/res-preserving-sharing-data.aspx>
- Research Institute for Nature and Forest: Open data policy
<https://www.vlaanderen.be/inbo/en-GB/publications/open-data-policy-of-the-inbo>
- Francis Crick Institute: Scientific research data management policy
https://www.crick.ac.uk/sites/default/files/2018-07/crick_scientific_research_data_management_policy_0.pdf

Data Repositories

This list is subject to regular review by the UK DRI Informatics Steering Committee and associated Data Working Groups. Any changes to the list will be announced on Slack and an up-to-date list will be maintained on [UK DRI Portal](#).

General

- [Registry of Research Data Repositories \(re3data\)](#): a global registry of research data repositories.
- [FAIRsharing](#): A curated resource on (meta)data standards, inter-related to databases and data policies.
- Where no appropriate domain-specific repository exists, data should be deposited in a general repository (eg [Zenodo](#), [Dryad](#)) that makes the data discoverable, reusable and citable.

Software. Source code should be made available under an open-source licence and deposited in [GitHub](#), [GitLab](#) or [Code Ocean](#), which will provide a persistent link. It is recommended that you archive your GitHub repositories on [Zenodo](#) to obtain a DOI that [make your code easily citable](#).

Domain-specific

- Microarray & nucleotide sequence. Data must be deposited in a [MIAME](#)- or [MINSEQE](#)-compliant format. Note also that the [International Nucleotide Sequence Database Collaboration](#) operates between [EMBL-EBI](#), [NCBI](#) and [DDBJ](#) covering raw reads, alignment and assemblies to functional annotation and contextual information relating to samples and experimental configurations. Data submitted to any of these organisations are shared.
 - [ArrayExpress](#): EMBL-EBI. Functional genomics data from microarray and sequencing platforms. Experiments are submitted directly via [Annotare](#) or imported from the [GEO](#) database. For high-throughput sequencing-based experiments the raw data is brokered to the [ENA](#) (see below), while the experiment descriptions and processed data are archived in ArrayExpress. [How to submit](#). **Recommended.**
 - [European Nucleotide Archive \(ENA\)](#): EMBL-EBI. Nucleotide sequencing data. Submitted data are validated by automated quality checking. Submitted data can be kept private and made public later. (Note that the change of status can take up to 15 days.) [How to submit](#). **Recommended.**
 - [Gene Expression Omnibus \(GEO\)](#): NCBI. Microarray, next-generation sequencing and other forms of high-throughput functional genomic data. Submitted data may remain private until made public. [How to submit](#).
 - [Sequence Read Archive \(SRA\)](#): NCBI. Raw sequence data from next generation sequencing technologies, including Illumina, 454, IonTorrent, Complete Genomics, PacBio and OxfordNanopores, and alignment information in the form of read placements on a reference sequence. [How to submit](#).
 - [Genbank](#): NCBI. Annotated genetic sequence database. [How to submit](#).
 - [European Genome-Phenome Archive \(EGA\)](#): EMBL-EBI. For the permanent archiving and *managed* access of personally identifiable genetic and phenotypic data. [How to submit](#).
- Protein.
 - [Proteomics IDentifications Database \(PRIDE\)](#): EMBL-EBI. Proteomics data, including mass spectrometry and protein interaction data. [How to submit](#), see also the [ProteomeXchange Consortium](#) for GUI-based submission tool. **Recommended.**
 - [Universal Protein Resource \(UniProt\)](#): EMBL-EBI. Including databases of protein sequences and functional information ([UniProtKB](#)), clustered sequences ([UniRef](#)), and an archive of sequence data in the public domain ([UniParc](#)).
 - [Protein Data Bank in Europe \(PDBe\)](#): EMBL-EBI. Tools and resources for studying structures of biological macromolecules and their relationships to sequence, function and disease. How to [use](#) and [submit](#).
 - [IntAct Molecular Interaction Database](#): EMBL-EBI. A molecular interaction database and analysis system. Interactions are derived from literature curation or [directly submitted](#).
 - [PeptideAtlas – PASSEL](#): SRM/MRM data. Part of the [ProteomeXchange Consortium](#).
 - [Protein Circular Dichroism Data Bank \(PCDDb\)](#): Circular dichroism (CD) and synchrotron radiation CD spectral data and associated metadata. Entries are linked, when appropriate, to primary sequence (UniProt) and structural (PDB) databases.



- [Chorus](#): An online platform for storing, analysing and sharing mass-spectrometric data.
- Metabolomics.
 - [MetaboLights](#): EMBL-EBI. A cross-species, cross-technique database for metabolomics experiments and derived information, covering metabolite structures as well as their reference spectra, biological roles, locations and concentrations. [How to submit](#). **Recommended**.
- [Image Data Resource \(IDR\)](#): A repository for high-quality bio-image data, including cellular imaging.
- [BioStudies](#): EMBL-EBI. Descriptions of biological studies, links to data from these studies in other databases at EMBL-EBI or outside, as well as data that do not fit in the structured EMBL-EBI archives. [How to submit](#).
- [AMP-AD Knowledge Portal](#): An NIH-designated repository for multi-omics data from human samples, cell-based and animal models, analysis results, analytical methodology and research tools generated by multiple National Institute of Aging supported AD research programs and consortia. (Hosted on [Synapse](#)).
- [DPUK Data Portal](#): Population cohort data, including imaging and genetics.
- [Biological General Repository for Interaction Datasets \(BioGRID\)](#): a curated database of protein-protein interactions, genetic interactions, chemical interactions, and post-translational modifications from major model organism species. [How to submit](#).
- [Systems Biology: Model repositories](#): links to resources where biological models generated using systems approaches can be deposited.
- [Expasy](#): The bioinformatics resource portal of the SIB Swiss Institute of Bioinformatics, providing access to over 160 databases and software tools, developed by SIB Groups and supporting research domains, from genomics, proteomics and structural biology, to evolution and phylogeny, systems biology and medical chemistry.

Version	Date	Author	Notes
0.0	4/11/2020	I-Chun Lin	Draft for ISC discussion.
0.1	26/1/2021	I-Chun Lin	Draft for ISC revision.
0.2	11/03/2021	I-Chun Lin	Revised after ISC meeting (5/3/2021) & by Caleb
1.0	20/04/2021	I-Chun Lin	Approved by SSOB for institute-wide adoption