

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

Document Title:	Data Management and Sharing Policy
Version	1.23
Document Review Date	September 2021
Document Owner	Wes Hinsley

Contents

1. [Introduction](#)
2. [Definitions](#)
3. [Aim](#)
4. [Policy](#)
5. [Data Catalogue <https://data.dide.ic.ac.uk/>](#)
6. [Roles and responsibilities](#)
7. [Related documents](#)
8. [Effective date](#)
9. [Review date](#)

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

1. INTRODUCTION

The research undertaken by the [MRC Centre for Global Infectious Disease Analysis](#) is highly interdisciplinary and focuses on the transmission, evolution and control of infectious diseases in human and animal populations. The Centre is committed to the highest standards in academic research, data management, attribution of appropriate metadata, and data sharing.

A data audit of Centre academics in 2016 revealed custodianship of 184 datasets, However of these, only 21 (11%) were also generated by those same academics. The audit revealed those datasets that are used by multiple groups within the Centre, and correspondingly highlighted the potential benefits that could arise from the standardisation of data extracts across the Centre.

This document provides a framework for both the sharing and the management of data specifically held by Centre staff. We have deployed an online [Centre Data Catalogue](#), using the open-source [CKAN](#) platform, that primarily serves as a data catalogue function but can also be used as a data repository for unconditionally publicly shareable datasets. The metadata on every dataset held (including owner contact details) by the Centre are provided through this platform.

PRINCIPAL INVESTIGATORS PLEASE NOTE

While this policy document should be referred to in individual project-level grant applications, its existence does not negate the obligation upon Centre Principal Investigators (PIs) to complete project-application-specific Data Management Plans (DMPs), where requested by Funders. To this end, PIs are advised to make reference to this document in conjunction with the use of [DMPonline](#), a free tool created by the Digital Curation Centre (DCC) that will provide a funder-appropriate DMP template.

2. DEFINITIONS

- 2.1. *Research data*: Data that are collected, observed, generated, created, or obtained from commercial, government or other sources, for subsequent analysis and synthesis to produce original research results. For the purposes of this policy, this is limited to digital data.
- 2.2. *Digital Object Identifier (DOI)*: A reference that enables permanent identification and location of digital objects, such as datasets, software, and journal articles. This means that reference to a dataset elsewhere (e.g. in the journal article it supports or Imperial College's catalogue) is not going to break or become outdated as a web address might do.

3. AIM

The aim of this document is to provide a framework within the Centre that enables the academic, research, technical and research support services staff to adhere to the rules and the mandates of funders; to comply with the broader Imperial Research Data Management Policy; and to outline the Centre-specific policies and procedures for data management and data sharing.

This policy document is fully complementary to the stated aims and content of Imperial's overarching [Research Data Management Policy](#), differing only to add local-level specificity in relation to management of datasets, dataset sharing, data storage, and data security protocols.

4. POLICY

The policy outlined below applies to research data that have been generated through research undertaken by Centre staff and students solely or in collaboration with others. This includes, but is not limited to, research sponsored by grants or contracts awarded to Centre staff where they are a partner on awards held by other institutions.

4.1. General Principles

- 4.1.1. The MRC Centre is in full agreement with the [Principles and Guidelines for Access to Research Data from Public Funding](#) developed by the Organisation for Economic Cooperation and

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

Development (OECD), and the [Research Council's Common Principles on Data Policy](#), and recognises the 'public good' inherent in the prompt sharing of research data wherever possible.

4.1.2. Research data held within Centre must be maintained and preserved as identifiable, retrievable, and reusable as a part of the overall research project cycle.

4.1.3. Where specific funder policy does not otherwise overrule, Centre academics retain the choice of where to deposit their data.

Centre PIs may select the licence under which they publish their data provided the chosen licence permits sharing and re-use in accordance with the appropriate funder's terms and conditions. Management of research data obtained under licence must also comply with relevant licence agreements. The terms and conditions of any licence take precedence over this policy.

4.2. Storage of Datasets and Software

4.2.1. Research data must be accompanied by appropriate metadata to facilitate identification and re-use.

4.2.2. During the course of a research project, Centre PIs must ensure their own and their group's data are stored and described in ways that minimise the risk of loss and maximise the likelihood of future discoverability and re-use.

4.2.3. Cloud storage of data is not prohibited but would not usually be considered best practice unless strong encryption is used. Minimum standards on data safety set by Imperial College must be followed.

4.2.4. Where software is developed as a part of a research project, Centre PIs are encouraged to archive the specific version of the software used to generate or analyse research project data that results in publication. [Imperial best practice](#) is to use a service such as [Zenodo](#) for storage, thus allowing the stored software to be issued with a DOI for the software at time of deposit.

4.3. Data catalogue and data sharing

4.3.1. Under the [RCUK Common Principles on Data Policy](#), a statement on data access is obligatory in relation to all research publications produced by Centre authors. To comply with these principles, and Imperial's broader policy on metadata provision, Centre PIs must create a dataset record on the [Centre Data Catalogue](#) for each dataset generated by their own research or received from collaborators (including those that are already in the public domain). Where applicable, each dataset will contain metadata on data type, disease type, age range, geography, date range, and ownership information.

4.3.2. It is also recommended that PIs record their use of commonly used publicly accessible demographic and geographic datasets, to minimise repetition of post-processing in the Centre and allow reference versions to be curated and internally shared.

4.3.3. For each dataset held, the corresponding PIs must specify whether it is able to be shared, both internally between Centre staff, and externally. In each case, there are four possible sharing conditions.

Yes - unconditional	The data can be shared unconditionally.
Yes - after specified date	The data can be shared unconditionally after a specified date.
Yes - conditional	Data may or may not be shareable and the data owner's approval must be sought.
No - unconditional	Data cannot be shared.

4.3.4. It is Centre policy that, wherever possible, data allowing reproduction of published research results shall be made publicly available with every publication, either as supporting information or on a separate data repository. In the latter case, the published data should be issued with an appropriate DOI. All such datasets should be listed on the Centre data catalogue.

4.3.5. Centre PIs are entitled to publish their findings first, and correspondingly may choose to categorise their dataset availability as '*yes – after specified date*' prior to publication. In such cases, and when

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

there are no other reasons for non-sharing of data, then the latest possible date for this categorisation will be **two years after the end of the grant that funded the data collection**. After this date, the data availability categorisation will revert to 'yes – unconditional'.

- 4.3.6. Where datasets are unlikely to be published in the near-term but are of urgent wider public value, sharing prior to publication is strongly encouraged.
- 4.3.7. Where data protection or data owner confidentiality prevent sharing datasets upon publication, each publication should provide a contact email address for data access inquiries. Such datasets should also be publicly listed on the [Centre Data Catalogue](#), with the email address of a Centre PI for data access queries.
- 4.3.8. Centre PIs are expected to provide up-to-date metadata to the data catalogue at point of receipt of a dataset, or generation of a dataset. An annual audit and update process will be undertaken to ensure the data catalogue remains accurate and current.
- 4.3.9. Research data held within Centre must also comply with the policies of relevant grant funders with regards to data sharing. Where such policies conflict with Imperial College or department plans, the policies of funders will take precedence.

4.4. Data Privacy and Ethics

- 4.4.1. The overwhelming majority of Centre-held data are anonymised and/or aggregated. In the rare cases where data may contain identifying information, the PI should take all appropriate steps to ensure data protection regulations are met. Such datasets should not be shared without approval of the data owners and suitable ethical approval. Where uncertainty remains, PIs should discuss specific cases with the School of Public Health (SPH) Data Protection officer (currently Eric Johnson; eric.johnson@imperial.ac.uk).
- 4.4.2. The advent and implementation of the May 2018 General Data Protection Regulation (GDPR) will expand the definition of personal data. The Regulation will require organisations not only to legally comply but also to be able to evidence such compliance clearly (e.g. through records and procedures). It is expected that the GDPR will be strictly enforced by the Information Commissioner's Office, with significant penalties for any breaches. It is anticipated that the Centre's compliance with GDPR will be operationalised through the [Centre Data Catalogue](#). A working group exists at Imperial and the Centre will work with it to ensure consistency with [Imperial policy and procedure](#) relative to the new regulations is observed.

4.5. Legacy Obligations

- 4.5.1. PIs will ensure that shareable research data are made publicly available for 10 years after publication or for the minimum period required by the funder if longer than 10 years.
- 4.5.2. If a PI leaves Imperial College or circumstances (such as death, incapacitation, or leaving academia) mean the PI is no longer able to fulfil their responsibilities then;
 - 4.5.2.1. If there is another Imperial College Co-investigator, they inherit the responsibilities of the PI;
 - 4.5.2.2. Otherwise, if the PI and the award move to another institution, the responsibilities follow the PI;
 - 4.5.2.3. Otherwise, whosoever is appointed by the Centre to take over the research projects of the PI will inherit their responsibilities as a data custodian.

4.6. Funding

- 4.6.1. Appropriate funding should be requested at application stage, on all proposals where data management costs are an eligible cost item allowable by the funder

5. DATA CATALOGUE (<https://data.dide.ic.ac.uk/>)

- 5.1.1. The [Centre Data Catalogue](#) acts as a data catalogue for **all** PI-held datasets currently held within the Centre. It contains metadata information on every dataset.

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

- 5.1.2. Each dataset in the catalogue must have the following minimal set of metadata: (a) Unique identifier (dataset name); (b) Data custodian within the Centre, with contact email address; (c) Data sharing status (i.e. not shareable, conditionally shareable, shareable).
- 5.1.3. All publicly shareable datasets listed on the data catalogue will have a URL to a downloadable copy of the data listed, either as supplementary information in a publication or as a DOI to Zenodo or another recognised data repository.
- 5.1.4. The [Centre Data Catalogue](#) is publicly accessible, though only publicly sharable datasets are guaranteed to be listed without Centre logon credentials.
- 5.1.5. A user account is neither required to search Centre Data Catalogue entries fully, nor to download any Centre (unconditionally shareable) datasets that have been stored using the [Centre Data Catalogue](#).
- 5.1.6. At any given time there will be at least one administrator profile (*authority to override or amend content and to create new editor accounts*), and a fixed number of editor accounts. The current administrator is Wes Hinsley (w.hinsley@imperial.ac.uk).
- 5.1.7. All Centre PIs will be eligible for an editor profile and will be able to add or amend their corresponding entries in the catalog.
 - 5.1.7.1. Editor profiles will also be issued to Centre research support staff who may be sanctioned by PIs to act on their behalf to add and amend catalogue entries and corresponding metadata. In such cases, it will be the responsibility of the PI to interact with and update the editor on a regular basis.
- 5.1.8. At minimum, editors will take part in an annual audit process with their corresponding PIs to ensure that all held datasets and associated metadata are represented on the [Centre Data Catalogue](#).

6. ROLES AND RESPONSIBILITIES

6.1. MRC Centre for Global Infectious Disease Analysis

- 6.1.1. PIs have overall responsibility for the effective management of research data generated within or obtained for their research, including by their research group members.
- 6.1.2. This responsibility extends from the application stage, with the provision of funder-compliant DMPs, through to the post-project stage where data must be appropriately stored.
- 6.1.3. Appropriate metadata for each Department-held dataset will be provided by the PI, or by members of their group, to the [Centre Data Catalogue](#) thus facilitating clear identification and eligibility for sharing.
- 6.1.4. To ensure that the extent and content of the dataset metadata that are represented within the catalogue remains accurate a recurring regular audit process of PI-held data, will be undertaken by Centre research support staff. The minimum frequency for this shall be annual.

6.2. Faculty and JRO

- 6.2.1. Research Offices and Contracts (JRO) are responsible for ensuring that all stakeholders understand their obligations, through the clear communication of funder and Imperial College requirements.
- 6.2.2. As a part of the routine approval of applications, JRO will ensure that all research grants and contract applications include a DMP where required by funders, and that the DMP is attached to the relevant records in the appropriate Imperial College management system.

6.3. Library

- 6.3.1. The Imperial College Library has a responsibility to provide training and guidance on all matters relating to data management, including the writing of DMPs, copyright, and licences.

6.4. Information and Communication Technologies service (ICT)

DATA MANAGEMENT AND SHARING POLICY

MRC Centre for Global Infectious Disease Analysis (MRC GIDA)

6.4.1. ICT has a responsibility within Imperial College to provide technical advice and support relating to data management, including data storage, backup and archiving. Clarification on roles and responsibilities can be found via ICT webpages online.

7 RELATED DOCUMENTS

7.1. Imperial College London

- Research Data Management Policy (2016) - <http://www.imperial.ac.uk/media/imperial-college/research-and-innovation/research-office/public/Imperial-College-RDM-Policy-%5Bpdf%5D.pdf>
- Research Office – Research Data Management - <http://www.imperial.ac.uk/research-and-innovation/research-office/research-outcomes-outputs-and-impact/research-data-management/>

7.2. Funder Policies

- RCUK – Principles on Data Policy - <http://www.rcuk.ac.uk/research/datapolicy/>
- MRC – Data Sharing Policy - <https://www.mrc.ac.uk/documents/pdf/mrc-data-sharing-policy/>
- NIHR – Data Sharing - <https://www.journalslibrary.nihr.ac.uk/information-for-authors/data-sharing/>
- BBSRC – Data Sharing Policy - <http://www.bbsrc.ac.uk/about/policies-standards/data-sharing-policy/>
- BMGF – Data Access Principles - <https://docs.gatesfoundation.org/documents/faq.pdf>
- Wellcome – Policy on Data Management and Sharing - <https://wellcome.ac.uk/funding/managing-grant/policy-data-management-and-sharing>
- NIH – Data Sharing Policy - https://grants.nih.gov/grants/policy/data_sharing/index.htm
- H2020 – Guidelines on Data Management - http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

8 EFFECTIVE DATE

This policy is effective from July 2017, and was reviewed in September 2021

9 REVIEW DATE

The policy will be reviewed in September 2022