

DATA SHARING GUIDANCE FOR CRUK RESEARCHERS

INITIATIVES AND REPOSITORIES TO SUPPORT POPULATION RESEARCH COMMITTEE RESEARCHERS WITH DATA MANAGEMENT AND SHARING

CONTENTS

1. Introduction	1
2. Generalist data sharing initiatives and repositories	2
2.1. Initiatives.....	2
2.2. Repositories	2
3. Initiatives and repositories specific to data sharing in Population Research Committee disciplines	3
3.1. Initiatives.....	3
3.2. Repositories	3

1. INTRODUCTION

In the following pages, we highlight some key initiatives and repositories which may provide tools and solutions for data sharing in fields within the remit of CRUK's [Population Research Committee](#) (PRC). Some of these are not discipline specific and are described as “generalist” ([Section 2](#)); others are discipline-specific, relevant to the kinds of epidemiology and behavioural research funded through PRC. It is the responsibility of the investigators to ensure that any repositories/standards/tools they intend to use are appropriate for the nature of the research envisaged.

CONTACT US

We would be grateful for any comments or suggestions to help improve this guidance. Please get in touch with jamie.enoch@cancer.org.uk with any feedback.

2. GENERALIST DATA SHARING INITIATIVES AND REPOSITORIES

2.1. INITIATIVES

Listed below are generic initiatives supporting data sharing in science and health research generally, which may provide useful tools, resources and methods you can factor into your data sharing plan.



[DataCite](#) provides advice on creating digital object identifiers (DOIs) for datasets. It also offers a range of services including a [Metadata search](#), which allows researchers to locate individual datasets through access to the relevant metadata, and a [search tool](#) to discover appropriate research data repositories worldwide through [Re3Data](#).



The [Digital Curation Centre](#) provides expert advice and practical help to researchers to store, manage, protect and share digital research data. It maintains a range of resources including [How-to Guides and checklists](#), [case studies in research data management](#), and [training programmes](#) for researchers and data custodians in research data management and sharing.



The [Expert Advisory Group on Data Access](#) convenes leading researchers on issues of data access and sharing who advise the funders on technical and legal issues in data sharing. It has produced a number of reports which advise on good practice in data sharing policy and governance.



[GigaScience](#) is an open-access open-data journal for ‘big data studies’ from across the life/biomedical sciences, in collaboration with BioMed Central. In many ways it is half journal and half repository. It links a standard manuscript publication with an extensive database, the [GigaDB](#), which hosts associated data and provides data analysis tools and cloud-computing resources.



[Nature Scientific Data](#) publishes peer-reviewed articles known as data descriptors, which focus on helping others reuse data by describing the dataset with structured, machine readable information. The articles can be descriptions of datasets of any size, and can link to datasets underpinning published research or describe standalone datasets. The journal mandates the release of datasets accompanying manuscripts, and links to datasets hosted on third-party repositories.



The [Research Data Alliance](#) works to build the social and technical bridges to facilitate data sharing and re-use. Its constituent [working groups](#) aim to tackle fundamental issues in data sharing; for example, [one working group](#) is seeking to rationalise databases, standards and funder policies in the biomedical sciences by working with the BioSharing platform.

2.2. REPOSITORIES

While good practice in clinical trial data sharing generally favours a managed data access approach, these generalist repositories may be appropriate for the storage of de-identified or aggregated data.

REPOSITORY	TYPES OF DATA RESEARCHERS MAY SUBMIT
Dryad Digital Repository	Data underlying scientific or medical publications
Figshare	Various data types, including figures, datasets and images
GigaDB	Data and tools for <i>GigaScience</i> and other articles
Zenodo	Research outputs from all fields of science

3. INITIATIVES AND REPOSITORIES SPECIFIC TO DATA SHARING IN POPULATION RESEARCH COMMITTEE DISCIPLINES

3.1. INITIATIVES

In the population research space, there are a number of initiatives working to facilitate data sharing. Those which support data sharing in epidemiology are generally seeking to provide improved infrastructure for integrating health records with research data; while those relevant to the behavioural science remit of CRUK's Population Research Committee are often supporting data in social sciences sharing more generally.



The [Administrative Data Research Network](#) is a UK-wide public sector partnership which offers a bespoke service for researchers carrying out economic and social research using administrative data. It does not hold administrative data itself, but works with government agencies to make the data available to researchers who carry out the work in secure environments provided by the Administrative Data Research Centres.



[BioSHaRE](#) is an EU information portal, providing a [catalogue of tools and services](#) which can facilitate sharing of data from key European cohort studies, such as the National Child Development Survey in the UK.

Cohort Directory

More for discovering than sharing data, the [cross funder cohort directory](#) of large cohorts funded by CRUK, MRC and other agencies provides a searchable, high level overview of cohort profiles and the types of variables and data collected.



The [Data Documentation Initiative](#) is a collaborative, international initiative aiming to set [specifications](#) for metadata describing statistics, behavioural or social science datasets throughout every phase of the data lifecycle. DDI also maintains a number of resources including [working papers](#) on best practices and use cases, and open source [tools](#) for editing metadata.



The [Farr Institute](#) of Health Informatics Research, through its four nodes across the UK, aims to deliver research linking e-health data with other forms of data, and build capability in health informatics research. They are also helping advance [e-infrastructures](#) in health informatics research, forming digital laboratories for larger scale research across the Centres and adopting a coordinated approach to the establishment of safe havens for data.



The [Public Population Project in Genomics and Society](#) (P3G) aims to provide resources to encourage collaboration, knowledge transfer and sharing of data between researchers and biobanks. Tools include a [Biobank Toolkit](#), a one-stop location for access to biobanking websites, software, protocols and systems.



The [UK Data Archive](#) and [UK Data Service](#) are key resources for storing, sharing and discovering social science and longitudinal study data, as well as data collected by UK public bodies. The UK Data Archive also provides [advice and training support](#) for every stage of the "[research data lifecycle](#)", and [guidance on legal and ethical issues](#) in research data management.

3.2. REPOSITORIES

These suggested repositories could all be appropriate for behavioural research data. There are currently no large-scale third-party repositories specifically for epidemiological research data, although for certain studies, a repository such as the UK Data Service may be appropriate. For epidemiological research data, a managed access approach with data stored in an institutional repository and made publicly discoverable via publications and project websites may also be feasible.

REPOSITORY

TYPES OF DATA RESEARCHERS MAY SUBMIT

[Harvard Dataverse Network](#)

Multidisciplinary social science data, either publicly available or selectively restricted

[Inter-university Consortium for Political and Social Research \(ICPSR\)](#)

Data from social and behavioural research

[Qualitative Data Repository](#)

Data used in qualitative and multi-method social inquiry

[UK Data Service](#)

Various types of social science data and data from cohort studies