DATA SHARING GUIDANCE

RESEARCH SUPPORT TEAM, UNIVERSITY OF NOTTINGHAM LIBRARIES

At the University of Nottingham, we want to encourage researchers of all disciplines to share the data underlying their research outputs. If your data is useful to you, it may well be useful to others, helping to drive forward research within your field.

This guide gives you some general advice on data sharing, but also provided specific guidance on using the UoN Data Repository to share your data.

For further enquiries and advice, contact the Libraries Research Support Team library-researchsupport@nottingham.ac.uk

FAIR DATA SHARING

Best data sharing practice subscribes to FAIR Principles (FAIR Principles - GO FAIR (go-fair.org)) which stand for:

Findability

The data, and associated metadata, should be easy to find for both researchers and computer systems. Unless the data can be found, it is unlikely to be re-used. Ensuring your metadata is clear, describes the data well for a non-expert user and includes keywords that other researchers are likely to use, is a great way to make sure that others can find your data.

Accessibility

Once the dataset has been found, it should be clear how a potential user gains access to the data. If you are sharing using a CC-BY license, there will be a clear download button, but if the potential user needs to sign a contract or contact the author before gaining access, this should be stated clearly.

Interoperability

Data should be shared in such a way as to ensure that it can be used in a wide variety of systems, and other datasets. Ways to ensure this are to use standard and open-source file types, and to ensure that you share data in the formats that are most commonly used in your field.

Re-use

To ensure that data has the most impact possible, we need to help others to re-use the data. Clear metadata and descriptions, commenting of software, and open licences all help to ensure that your data can be reused by other researchers. Reuse of data leads to data citations, or even new collaborations.

PREPARATION DECISIONS FOR SHARING DATA

OVERVIEW

1. **WHAT DATA SHOULD YOU KEEP?**
2. **IS IT SENSITIVE OR CAN IT BE SHARED OPENLY?**
3. **WHERE IS THE BEST PLACE FOR THE DATA?**
4. **HOW LONG SHOULD THE DATA BE KEPT FOR?**

DETAILS

1. Think about all your data. Any data which falls under one or more of the following categories you should archive, and share openly if possible:
    a. Data which supports a research output (journal article, monograph etc.)
    b. Data which would be difficult for someone to re-collect e.g. data taken during rare conditions or data collected during important events
    c. Data which is likely to be of interest to other researchers in their work e.g. biological sequence data or standardised measurements

2. For each type of data you wish to collect, think about whether the data is sensitive (personally or commercially) and how open the data can be:
    a. Not sensitive; Can be shared under a CC-BY licence
    b. Commercially important but not sensitive: Consider using a CC-BY-NC licence
    c. Not sensitive, but you're still working with it; Place a 1-3 year embargo, transitioning to a CC-BY licence
    d. Sensitive, but could be shared with other researchers; Place a permanent embargo and require a data sharing agreement
    e. Sensitive, cannot be shared; Place a permanent embargo.
    f. Extremely sensitive; Consider using the UoN dark archive where not even the metadata (data description) is shared. To access this service, email libraryresearchsupport@nottingham.ac.uk

3. Think about where the best place is for your data. UoN's order of preference is:
    a. A repository mandated or suggested by your research funder e.g. UK Data Archive or one of the NERC data centres
    b. A subject specific repository e.g. the Worldwide Protein Data Bank or the UK Solar System Data Centre. Options can be found at https://www.nature.com/sdata/policies/repositories and https://www.re3data.org/
    c. The University of Nottingham Data Repository https://rdmc.nottingham.ac.uk/
    d. A general data repository e.g. Figshare, Zenodo or Dryad

4. How long will the data need to be stored for?
    a. Does your research funder have a requirement for how long data should be kept?
    b. Have your research participants given consent for their data to be archived for a specific length of time?
    c. The University of Nottingham retention schedule states that research data should be kept for a minimum of 7-year years, and usually up to 25 years.

It is important to note here that where a DOI is issued but the data is later deleted, a tombstone page will be created which gives the full data citation and reasons for the unavailability of the data.

PREPARATION ACTIONS FOR SHARING DATA

OVERVIEW

1. **ENSURE YOUR DATA IS IN A FORMAT WHICH PROMOTES RE-USE AND INTEROPERABILITY**
2. **ENSURE ANY SOFTWARE IS COMMENTED AND READ-ME FILES ARE AVAILABLE**
3. **PREPARE ANY NECESSARY METADATA AND DOCUMENTATION FILES**

DETAILS

1. Think about what format your data is in. Is this the easiest way for other researchers to access the data?
    a. Is the format something that is commonly used in your field, or does it require extra explanation to allow another researcher to use it?
    b. Are column headings or variable names understandable, or do you need to include a key?
    c. Can the data be accessed using free software, or does it need proprietary software? If the proprietary software is commonly available in universities or other organisations in your field (e.g. Matlab) this may be fine, but consider what will happen if the company folds or stops making the software you are using.
    d. To aid 'Interoperability' and 'Reusability' (from the FAIR guidelines) we would recommend you either use very commonly used formats (e.g. .docx, .xlsx) or open formats (e.g. .txt, .csv).

2. Is any software that you're sharing easily understandable by other researchers?
    a. Do you have a read me file including: The purpose of the code or software, how it fits with any other scripts, what inputs are needed and what format these are in, what output are created and what format these are in, and any assumptions or options that are either coded in or should be altered as necessary by the user?
    b. Have comments been included in your code highlighting any main functions?

3. Metadata is used to explain your data, make sure that the variables are meaningful and clear, and to ensure that researchers (including you) can come to the data in the future and use it without having to ask the authors for further clarification and assistance.
    a. Does the repository you plan to use have a formal metadata schema? Filling out the fields available to you in the fullest way possible will make it clear what your data is even before it has been downloaded.
    b. Are there additional files you need to include? e.g. readme files for software, or keys explaining your variables.
    c. Check out the UK Data Service advise on providing clear documentation https://ukdataservice.ac.uk/learning-hub/research-data-management/#document-your-data

DEPOSITING IN UON DATA REPOSITORY

OVERVIEW

- **FOLLOW THROUGH THE SUBMISSION FORM AT RDMC.NOTTINGHAM.AC.UK. * INDICATE MANDATORY FIELDS.**
- **PROVIDE AS MUCH INFORMATION ABOUT YOUR DATA AS YOU CAN.**

DETAILS

The following table shows you all the available metadata fields alongside some guidance and tips on how to get the most out of using these fields. Some of these fields are mandatory and are marked by an asterisk (*).

Field Title	Description	Guidance and Tips
Title *	The title is one of the most important metadata fields, this is what will help other researchers decide if the data deposit is going to be useful to them.	The title of the paper is already linked to the associated data, so using "data from [paper title]" is a wasted opportunity to give the reader more information What types of data are in the file? What is the purpose of the data? Useful: Beak length measurements of birds in Scottish Highlands – Field Season Aug 2021 Less useful: Data from: Evolutionary development of bird beaks
Alternative title	A non-mandatory field allowing you to put in an additional title for your dataset.	If you want to include the title of the paper this is best field to add to.
Resource Languages *	The language(s) used in the dataset.	You can select more than one language for each data deposit. However, if you have data in more than one language you might considered separating these using a folder structure, to make it easier for others to find the right data.
Description:		
Creators:	This field should include the person(s) responsible for the creation of data.	*The lookup button does not function properly. This will be fixed in future versions* This is the first field where you can add people involved in the creation of the data. This is the field for the primary creator/owner of the data. The contributor's field can be used to include more collaborators and researchers.
Schools/ Departments *	Include all UoN schools and departments	Use the division look up under the field to navigate to the right schools and departments.

Research Institute and Centres	Include any official UoN Research Institutes and Centres	Use the division look up under the field to navigate.
Contributors	Add the names of all contributors.	*The lookup button does not function properly. This will be fixed in future versions* Treat this field like the author list in a publication and remember to include here the names of contributors who are not students/staff at UoN.
Corporate contributors	Include corporate partners who helped to create or analyse the data.	This is a free text field so it is important to check that you've used the official name of any partners and that this is spelt correctly, otherwise data may not be found when searching.
Funder (lookup) *	Add the funder of the project	Use the lookup option below the field. Where a funder is not available in the list, or the project is internally funded select 'other' and fill in the free text option.
Funder (free text)	Write in your funder if it isn't already included in the lookup list in the previous field.	Check that your funder isn't included in the lookup list first. For internally funded projects, or self-funded PhDs, use "University of Nottingham".
Grant number	Include grant numbers for externally funded projects	Leave this field blank for internal or unfunded projects
Data Type	Include all data types present in the dataset	This is a free text field allowing you to describe your data in whichever way is most appropriate.
Collection dates	Add dates when any data collection took place	To aid clarity around collection dates use the ISO 8601 format: YYYY-MM-DD or YYYY-MM or YYYY.
Temporal extent	Add timeframe of data collection and events specific to your dataset	If your data refers to events that happened at a particular moment in time, include that information here. This may particularly be relevant when the collection dates are not the same as the event about which data has been gathered.
Geographic coverage	Add countries, regions or towns related to your dataset	If your data is collected from a particular location, you can include the names of any areas. For areas which are not well described using town/region/country names, use the coordinates field below.
Geographic coordinates	Add co-ordinates of the bounding box for your data	If your data is collected from a particular location, and you have the latitude and longitude coordinates to describe this area, you can use this field to do so.

Data collection method	Include brief details about data collection methods, including any particular equipment used or approaches taken.	If your data management plan is up to date then you can copy and paste a lot of this information from there.
Parent dataset	Add the DOI or URL of any dataset which this data is based on or a subset of.	This is a good way of ensuring that datasets are linked together and datasets which aren't independent are clearly marked, the parent dataset can be in the UoN data repository but can also link to an outside dataset.
Keywords *	Include keywords which will help others find your data	The keywords will be indexed by search engines such as Google. Alongside your title and other metadata this will help others find your data. Think about what words you would use to find a dataset like this. Think about what other uses the data could be put to. Think about non-academic terms people might use to search for data. Are there any commonly used synonyms in your field?
Subject classification - JACS	A formal classification scheme to help your dataset be found.	Choose a subject which describes your data. This may help other researchers find the data when browsing or searching.
Subject classification - LC	A formal classification scheme to help your dataset be found.	Choose a subject which describes your data. This may help other researchers find the data when browsing or searching.
Publication Date *	Add the date when the dataset will be made publicly available.	If the dataset is to be made immediately open add the current data as the publication date. If you are setting an embargo, set this to the date when the embargo will expire.
Publisher *	Add the institution publishing the data.	If you are depositing data files in this repository the publisher is The University of Nottingham. If this is a metadata only record, include the name of the organisation who is archiving and sharing the data.
Associated publication DOI	Add DOIs of any publications using this data.	More than one DOI can be added to this field, so include all relevant publications as it provides greater linking between your research outputs. Future publications can be linked to this data deposit by emailing

		library-researchsupport@nottingham.ac.uk
Related resources	Website URLs or links to documents or datasets without a DOI can be added here.	For anything with a DOI add it to the relevant dataset or publication field, however project websites, policy documents or working papers can be linked to from this field.
Parent project Project name or identifier.	If this dataset comes from a larger project, give the name or identifier of the project.	This can help group together datasets which are related and can help to point towards other relevant resources.
Legal and ethical issues	Description of any legal or ethical issues which have had to be overcome to share the data, or which prevent the open sharing of your data.	This section can be primarily taken from the ethics section of your data management plan. It should cover issues such as consent, anonymisation and ownership of the data.
Additional information	This field can be used to include any information which has not already been covered.	If there's something important about your dataset that hasn't been covered anywhere else, now's your chance!
Rights owner	Add the name of the data asset owner. This is likely to be the PI of the project or research group, or in the case of PhD data it may be the PGR student.	Although UoN owns the copyright of all data created at the University, this field gives the name of the responsible party when dealing with the data.
Contact email address	The contact email address that Libraries can use.	This email address will not be publicly available but will be used in the future if Libraries need to contact you about your data. In particular, this will be used at the end of the 7-year period to ask whether data should be deleted or not. We recommend that you use a department or group email address where possible as this is less likely to need to be updated.

ADVERTISING, CITING AND LINKING TO YOUR DATASET

OVERVIEW

- **INCLUDE A DATA ACCESS STATEMENT IN YOUR JOURNAL ARTICLES**
- **INCLUDE A**
- **LINK TO THE DATA DOI FROM RIS OR ETHESIS**
- **ENSURE YOU CITE ANY DATASETS YOU USE**

DETAILS

The following table shows you

1. Include a Data Access Statement (often called a Data Availability Statement or DAS) in your journal articles.
    a. A DAS should include your DOI and any access conditions e.g. if the data is embargoed for 2 years or if a contract is needed to release the data.
    b. Your journal may have a particular section for a DAS, but if not, you can include it in the acknowledgements, or at the end of the main text of your article.

2. Link to the data deposit from other systems
    a. RIS contains a field which allows you to include the DOI of any related data sets (Go to your 'In Progress' output -> Library tab -> Use the Data Location URL field
    b. To add datasets to 'In Review' or 'Discoverable' outputs contact openaccess@nottingham.ac.uk

3. In a similar way to citing literature that you use in your work; you should also cite any datasets that you use. Datasets can be cited in a similar way to literature and most citation formats will have guidelines.
    a. An APA example is: **O'Donohue, W. (2017). Content analysis of undergraduate psychology textbooks (ICPSR 21600; Version V1) [Data set]. ICPSR. https://doi.org/10.3886/ICPSR36966.v1**
    b. Further details on citing datasets can be found at https://ukdataservice.ac.uk/learning-hub/new-to-using-data/#cite-data-correctly