



University of
Nottingham
UK | CHINA | MALAYSIA

UoN generic Data Management Plan guidance

Research Support Team

May 2019

UoN Data Management Plan

Write 2-3 pages A4

Data description

What data will you create?

Data Description: List the kinds of data or major datasets that you expect to create or use, noting its estimated volume/size, content, coverage or scope, and the **types** of the data (e.g. quantitative, qualitative, survey data, experimental measurements, models, images, audio-visual data, samples, documents, spreadsheets, databases, images, film, digital models, contents of an application, hardcopy records, physical collections or manuscripts, physical samples, or instrumental measurements etc.). Include the raw data arising directly from the research, the reduced data derived from it, and published data. Also, include non-digital data if relevant. Describe the **types**, **volume**, and **scope** of data that will be created and managed in the following terms: quantitative, qualitative, generated from surveys, clinical measurements, interviews, medical records, electronic health records, administrative records, genotypic data, images, tissue samples, etc.

Data volume: It is important to plan for the volume of data you plan to capture, process, and store. Think about the volume of primary and secondary data you will be producing as part of your research. This could be articulated in the anticipated (estimated) number of files you will generate, or the projected storage capacity you expect to require.

Data format: What file formats will you be making use of during your research? Do formats and software enable sharing and long-term validity of data? UKRI and other funders prefer “open” formats to “closed”, proprietary formats, whenever possible. Formats should also be considered to preserve the highest standard of data (so, tiff files would be preferable to compressed jpgs for storing images, for example), and for long-term archival use when possible (e.g. PDF/A, rather than generic PDF, is better suited to long-term archiving). Using propriety formats is acceptable if that is the most appropriate format for your research. State the format of the final dataset. Describe the file **formats** and **software** used, the number of records, databases, sweeps, repetitions... (in terms that are meaningful in your field).

continues overleaf

Data collection and/or generation

What are your methodologies for data collection / generation? How will you ensure data quality?

What data standards will you use?

Describe how you will collect or generate your data, and justify why new data collection is needed. Describe good practice and community standards for ensuring new data are of high quality and processing is well documented. How will consistency and quality of data collection / generation be controlled and documented? I.e., through processes of calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies?

Before embarking on a complex project to generate new data, you should **assess existing datasets** to ensure that nobody has already gathered this data, and that you are aware of similar relevant datasets that already exist. Provide **references**. [To locate potential datasets of use in your discipline, consult RE3Data.org]. You will need to justify creating new datasets if data relevant to your research is available. Consider how the data you plan to collect or create relates to existing datasets – identify **gaps** and **relationships** between currently available and required data.

If purchasing or reusing existing data sources, explain how you address issues such as cost and intellectual property. You should consider the licencing terms to access/re-use data, and what the terms of use are. Give an overview of any contracts or access/use agreements you will have to sign, and any restrictions on how the data might be used and shared.

Methodologies: Outline the **proposed methodologies** you will use to collect or generate data, including **software** and tools for data collection, and which community data **standards** (if any) will be used to ensure the data is of high quality. Provide information on the new data: volume, type, and content. Explain how you will ensure (control and document) data quality and consistency (**quality assurance**), such as standardised data capture or recording, digitisation, data entry validation, peer review of data, calibration, repeat samples or measurements, representation using controlled vocabularies, use of a uniform interview method to maintain consistency in qualitative data collection, etc. Indicate how the data will be organised during the project, for example, naming conventions, **version control**, and folder structures.

Ethics and privacy

Are there any ethical, commercial, or privacy issues that will affect the collection and storage of your data? Are you working with sensitive or personal data? Will you need to issue privacy notices?

Describe the legal and ethics procedures you will follow.

Ethics and informed consent

Ethical issues may affect **how you store data, who can see/use** it during collection and in future (data sharing), how long it is kept. You should show that you're aware of this and have planned accordingly. The University's Code of Research Conduct and Research Ethics provides guidance on what types of research will require ethical review. You are responsible for ensuring your handling of all this type of information is secure and complies with the law and associated research ethics procedures. Outline how you will obtain informed consent for the use of personal data as part of the [research ethics process](#) and to satisfy the duty of confidentiality towards the individual. The UK Data Service provides excellent [guidance in structuring consent forms](#) that respect legal obligations and facilitate [data archiving](#) and [publishing](#).

Data protection

If working with personal information you will have legal obligations with regard to data protection. You should provide a short statement around how you will comply with the GDPR/Data Protection Act 2018.

Sample text: "Personal data will be collected during this project, and the project has considered ethical and legal implications in its data storage, as well as appropriate security of personal data. All participants will agree to data collection and to long-term retention, archiving, and sharing of their anonymised data. Research will follow standard ethical procedures of the Faculty of Arts and the University of Nottingham. Specific aspects will be considered by the Faculty ethics committee as appropriate. In particular, the creation of data from interviews and focus groups will require ethical approval, including consent forms outlining the storage and use for research purposes of data, including access to those data by project researchers and other researchers, both during and after the life of the project. Participants will be informed that they can withdraw their participation at any stage during or after the observations. As we will be working with personal data we will ensure that we comply with the Data Protection Act 2018, including GDPR requirements. This will include providing research participants with the relevant privacy information and ensuring appropriate safeguards for the storage and handling of data are in place."

Data storage and security

Where and how will your data will be stored, backed-up, transferred, and secured during the active phase (short to medium term) of research?

Describe **where** and **how** your data will be **stored, secured, backed-up** and be **recovered**, and why these choices are appropriate (for security, data maintenance and recovery, and **appropriate costs**). Describe what devices will be used to capture, process, and store data. Describe back up procedures. When working in the field, describe how you will ensure data is encrypted and securely transferred to a University of Nottingham approved storage solution. You should also describe additional safeguards that you may utilise, such as the anonymization of personal and sensitive data.

Sample text: "We will use UoN-provided storage for our working data. UoN licenses **Microsoft OneDrive**, an ISO 27001 information security management compliant service that allows secure and controlled sharing of data amongst the research team. University of Nottingham OneDrive encrypts data both in transit and at rest and is approved against the University's Handling Restricted Data Policy. The service provides continual failover support. This service provides up to 5TB free-at-point-of-use, and as we do not anticipating generating more than 5TB we we will not require any additional costs for use of this service."

Sample text: "We will be utilising the **high-performance storage** option at the UoN, charged at £179 per TB/annum. We anticipate generating circa 10TB, and so we have accounted for this in the justifications of costs plan attached in the associated bid."

Data management, documentation and curation

What are your principles, systems, and major standards for data management and creation? What metadata and documentation will you keep?

Data documentation: Highlight what sort of documentation you intend to provide, including what software and processes/protocols will be required to understand, reuse, and validate the datasets. For example, descriptions of data that enable research data to be used by others outside of your own team. This may include documenting the methods used to generate the data, analytical and procedural information, capturing instrument metadata alongside data, documenting provenance of data and their coding, detailed descriptions for variables, records, etc.

Metadata: When depositing data, repositories will usually require descriptive metadata that describes the data in detail: typically a title, a brief description of the dataset, and the creator(s) of the data. Provide an overview of what metadata you will provide in order to facilitate discoverability and use of the data. There are various metadata standards which can help you to describe your data in a consistent way, and individual data repositories will advise on the relevant standards. You should provide as much information as you are able to. As an example of the kind of information requested, please see our [guide for depositing data](#) within the University of Nottingham's data repository.

Sample text: "We shall record and provide documentation via a README file that will describe:

- our research aims, objectives and hypotheses;
- our data collection methods, including the hardware and software used and any calibration carried out;
- a description of our data validation and quality assurance procedures undertaken;
- a description of the dataset structure and what each file relationships between files or versions of the dataset.

In addition, we shall provide templates of our interview sheets and consent forms."

Sample text: "On deposit of our data and associated documentation we shall provide catalogue metadata that will facilitate the discovery of our data. This shall conform with the DataCite metadata schema."

Data preservation

How will you ensure the long term storage and preservation of data?

Outline how you will ensure data is stored long-term: state where you will deposit your data (your nominated data centre or repository), and how long it will be preserved for. Some research funders and regulators specify how long data should be retained for. Where this is articulated, you should describe how you will meet this requirement (i.e. selecting an appropriate repository that guarantees preservation for a minimum period). In all other circumstances, the University expects data to be retained for a minimum of 7 years. Outline what data you will retain. Not all data will be suitable for retention, but some data will be mandated for retention. Identify any data repository (-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types. Most funders recommend that data underpinning

research outputs should be deposited in an established data repository that specialises in retention, preservation, and publication of data. When choosing a repository, you should consider:

- Are the repository's terms and conditions acceptable and do they align with funder requirements?
- Will support for access or curation be needed from the data centre that exceeds the norm, and has this been planned for/ costed?
- Where is your data going to be stored and does this meet legal or ethical obligations such as compliance with the GDPR?
- Will your dataset be given a permanent identifier, such as a DOI, so that the link to your data will stay the same even if the web URLs change?
- Is the repository used by researchers in your discipline? Does the repository already have a good reputation in your field and is it recommended by your funder or your journal?
- Does the repository allow you to describe your data sufficiently, so it is easy to find and easy to cite?
- Is the repository established and well-funded so that you can rely on it still preserving your data in 10 years' time?
- In your data management plan, you should state where your data will be deposited for long term preservation, and identify any costs associated with storing and making your research data available online.

If a data centre exists for your specific discipline, then it is usually recommended to choose this option. Some funders mandate deposit in a specific repository, for example, ESRC data must be deposited with the UK Data Service within 3 months of project end; NERC has a network of environmental data centres to provide a similar job for research it funds. See the NERC Data Centre or the ESRC-supported UK Data Service. You can check if a repository exists for your discipline by searching the re3data.org Registry of research data repositories

If there is no established data repository available for your discipline or funder, the University of Nottingham also has its own Research Data repository, which issues a DOI (Digital Object Identifier) for each dataset deposited so that it can be easily cited. The repository can be used to either deposit a closed (private) dataset, publish data openly, or under controlled access restrictions. The repository will provide access to your data in line with your funder's retention period.

Sample text: "All research data created by the project will be deposited in the UoN research data archive, <https://rdmc.nottingham.ac.uk/> For each published dataset, a DataCite DOI is issued

facilitating the ability to cite the data in associated research outputs. The UoN data archive is underpinned by commercial digital storage which is audited on a twice-yearly basis for compliance with the ISO 27001 standard. UoN will retain and preserve research data in line with [insert your funder's name] requirement for a minimum of [insert your funder's minimum retention requirement] years, but data will be retained for longer periods of time where it is of continual value to users. No cost has been charged to this project for data archiving as we anticipate that the amount of data generated for long-term retention will not exceed 50GB (the capacity provided free by the University)."

Sample text: "The UoN data archive provides up to 50GB of free capacity per research project and £94/TB per year thereafter. We anticipate 5TB data will need to be stored. This has been provided in our justification of costs."

Sample text: "All research data created by the project will be deposited in [name of the data repository] who will store and maintain the data on our behalf in line with our funder's requirements. There is no cost for this service."

Data sharing, publication and access

How will the data generated be shared and published?

Describe what data will be suitable for sharing and how this will be achieved (through a data repository, or a description of the data in a data journal). Indicate how potential new users (outside of your organisation) can find out about your data. Indicate whether the data will be deposited in and available from an identified community database, repository, archive or other infrastructure...

continues overleaf

established to curate and share data. If the data will be suitable for controlled access, identify who makes or will make the decision on whether to supply the data to a potential new user. Indicate whether external users are (will be) bound by data sharing agreements, setting out their main responsibilities. If the data will not be suitable for sharing, justify why this is the case. If you do not wish to share the data for a limited, defined period of exclusive use of data for primary research, show how this restriction on sharing is based on simple, clear principles. What are the timescale/dependencies for when data will be accessible to others outside of your team? Summarize the principles of your current/intended policy. Indicate whether your policy or approach to data sharing is (or will be) published on your study website (or by other means).

Outline the way(s) in which your data will be shared (e.g. DOI), and/or any justification for why this is not possible. **How** will data be shared, and what is its **value to others**? State the **timeline** for public release of data and the **mechanism** to achieve this. The preferred way to share data is by depositing it with data repositories who will either publish it openly online, or offer some form of controlled access. You can also publish a description of your dataset in a [data journal](#). You will need to include a data access statement in published outputs or final reports referencing your DOI, or, where and how data can be accessed.

Will the data be open, or will you charge to access it? Justify if charging.

How will you license the data?

How will you ensure that your data is **Findable, Accessible, Interoperable, and Re-usable**?

State when you intend to share or release your data, including any planned embargo period and the justification for this. Some communities have established timeframes for releasing data. These disciplinary guidelines, if they exist, should be referenced in the DMP. State any restrictions or delays to sharing, with planned actions to limit such restrictions. For most funders, a limited and defined period of exclusive use is acceptable (**embargoes**). This may be for reasons such as to avoid compromising research results by premature sharing, or to allow for the filing of patent applications.

It is also acknowledged that not all data can or should be 'open' and so publishing and sharing data may need to be restricted, for example:

- legal requirements – if your data includes personal information subject to the Data Protection Act 2018 and the General Data Protection Regulation (GDPR).

- ethical concerns – if your data includes personal or confidential data where no consent for sharing has been given.
- licence restrictions – if you are using data owned by third parties and don't have the rights to share.

If you believe your data will require controlled access state this here. Details of the managed access process should be provided, including if a Data Sharing Agreement will be completed for each data request. If you expect obstacles to sharing your data, explain which and the possible measures you can apply to overcome these. Make explicit mention of the planned procedures to handle consent for data sharing for data obtained from human participants, and/or how to anonymise data, to make sure that data can be made available and accessible for future scientific research.

Sample text: "Our dataset does not contain any personal or commercially sensitive information and thus will be shared via the University of Nottingham data archive [in line with the timelines mandated by my funder for the timely sharing of data]. There will be no need to update the data past the project period."

Sample text: "The anonymised data (transcripts and questionnaires) will be deposited in the University of Nottingham data archive once all data has been gathered, anonymised, and cleansed [in line with the timelines mandated by my funder for the timely deposit of data]. In order to minimise any further risk of disclosure, we shall control access requests to any data containing personal information to ensure the confidentiality of our research participants. Access agreements will be written on request in conjunction with the Research Contracts Team."

Sample text: "The existence, range, and nature of the project's original data will be publicised via references in published outputs by including relevant dataset DOIs, as well as via conference presentations and materials produced during the project."

Sample text: "Our data will be deposited in the University of Nottingham data archive. Access to this dataset will be embargoed for a period of 3 months in order for us to complete our analysis of the dataset."

Sample text: "There are no financial requirements for archiving and sharing data in the UoN data archive as the volume of data will not exceed 50GB."

continues overleaf

Roles and responsibilities

Who will be responsible for managing data, data security, data quality, and data security both during the award and post-award?

Outline who will be responsible for owning, storing, analysing and archiving your data within research teams at all partner institutions. Describe how these responsibilities will be regulated. Provide details on governance of access. Describe what resources will you require to deliver your data management plan. Describe how you will ensure quality across team members collecting and processing data.

Relevant policies

List the relevant institutional, departmental, or study policies.

Sample text: "We will ensure that our research aligns with the requirements of the University's Research Data Management Policy, Information Security Policy, Code of Research Conduct and Research Ethics."

"We will ensure that our research aligns with the requirements of the University's Research Data Management Policy, Information Security Policy, Code of Research Conduct and Research Ethics. As we are working with personal data, we will abide by the Handling Restricted Data Policy and Data Protection Policy. All third party commercial data or new data that may be suitable for commercial exploitation will be protected by the University's Intellectual Property policy."

Intellectual property

Who will own the copyright and IPR of any data that you will collect or create? Will you create a licence(s) for its use and reuse? If you are planning to use existing data as part of your research, do any copyright or other restrictions determine its use?

Copyright and IPR of new datasets: State **who will own the copyright and IPR** of any data that you will collect or create, along with the **licence(s) for its use and reuse**. An agreement concerning intellectual property ownership with a third party, such as a research funder or commercial sponsor may exist. For **multi-partner projects**, IPR ownership may be worth covering in a formal consortium or collaboration agreement. In the absence of such an agreement, the University of Nottingham asserts its ownership of rights in datasets generated using the institution's facilities.

Copyright and IPR for existing datasets: If you are planning to use existing data as part of your research, you should state the copyright or other restrictions which determine its use. If possible you should investigate these issues and attempt to gain copyright clearance so that your data can be shared at the end of your project. You should give full and appropriate acknowledgement, via citation, for any existing data that you use.

Sample text: "This project is being carried out in collaboration with an industrial partner. The University of Nottingham and our industrial partner will jointly own the data generated, as set out in the collaboration agreement. The intellectual property generated from this project will be fully exploited with help from the University of Nottingham's IP Commercialisation Office. The aim is to patent the final procedure and then publish the work in a research journal."

Sample text: "The intellectual property of the data generated will remain with the University of Nottingham."

Sample text: "The intellectual property of the newly created data will remain with the University of Nottingham. However, the use of archival material is done so under license from [insert repository/database] and thus the IP for this data remains solely with [insert repository/database]."

Budgeting

What are the costs or funding required for capturing, processing, storing, and archiving your data?

You should consider the costs or funding required for capturing, processing, storing, and archiving data. Where dedicated resources are required, these should be outlined and justified as part of your plan and included in your overall budget. Many funders will cover costs associated with data management and sharing provided they are incurred within the funding period. You may wish to consider: Accessing third party data - are there costs associated with accessing third party data, for example, licences, or commercial providers; people and skills - is additional specialist expertise (or training for existing staff) required?; storage and computation - will additional computational resources be needed in order to process, analyse or store your data?; archive – will there be any costs associated with preserving and/or sharing your research data, such as using a repository? Estimate the costs for making your data FAIR. Describe how you intend to cover these costs. Describe costs and potential value of long term preservation.