# CRUK DMP Template *(With Guidance Notes)*

### Admin Details

**Plan Name:** CRUK Template

**Principal Investigator / Researcher:**

**Funder:** CRUK

**Institution:** Royal College of Art

### 1. Data description

**Outline the volume, type, content and format of the final dataset**

***RCA/DCC guidance: Data description***

*Give a summary of the data you will collect or create, noting the content, coverage, format, volume and data type, e.g., tabular data, survey data, experimental measurements, models, software, audiovisual data, physical samples, etc.*

*Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse.*

*Indicate which data are of long-term value and should be shared and/or preserved.*

*If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.*

### 2. Data standards

**State the standards that will be utilised for data collection and management**

***RCA/DCC guidance: Data collection***

*Outline how the data will be collected and processed. This should cover relevant standards or methods, quality assurance and data organisation.*

*Indicate how the data will be organised during the project, mentioning, e.g., naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.*

*Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeat samples or measurements, standardised data capture, data entry validation, peer review of data or representation with controlled vocabularies.*

### 3. Metadata and documentation

**Outline the metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly**

***CRUK Guidance*** *For data sharing to be a success it is important that data are prepared in such a way that those using the dataset have a clear understanding of what the data mean so that they can be used appropriately. To enable this, applicants are encouraged to include with the dataset all the necessary information (metadata) describing the data and their format. This information should include such information as the methodology used to collect data, definitions of variables, units of measurement, any assumptions made, the format of the data, file type of the data etc. To support this researchers are strongly encouraged to utilise community standards to describe and structure data, (e.g. common terminology, minimum information guidelines and standard data exchange formats).*

***RCA/DCC guidance on Metadata & documentation***

*What metadata will be provided to help others identify and discover the data?*

*Researchers are strongly encouraged to use community metadata standards where these are in place. The Research Data Alliance offers a*[*Directory of Metadata Standards*](http://rd-alliance.github.io/metadata-directory/)*. Data repositories may also provide guidance about appropriate metadata standards.*

*Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, any assumptions made, the format and file type of the data and software used to collect and/or process the data.*

*Consider how you will capture this information and where it will be recorded, e.g., in a database with links to each item, in a ‘readme’ text file, in file headers, etc.*

### 4. Data sharing

**4.1 Outline the method used to share data**

***CRUK Guidance*** *The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. Data can be shared by any of the following methods:*

* ***Under the auspices of the Principal Investigator -****Investigators sharing under their own auspices may securely send data to a requestor, or upload the data to their institutional website. Investigators should consider using a data-sharing agreement (see below) to impose appropriate limitations on the secondary use of the data.*
* ***Through a third party -****Investigators can share their data by transferring it to a data archive facility to distribute more widely to the scientific community, to maintain documentation and meet reporting requirements. Data archives are particularly attractive for investigators concerned about managing a large volume of requests for data, vetting frivolous or inappropriate requests, or providing technical assistance for users seeking to help with analyses.*
* ***Using a data enclave -****Datasets that cannot be distributed to the general public due to confidentially concerns, or third-party licensing or use agreements that prohibit redistribution, can be accessed through a data enclave. A data enclave provides a controlled secure environment in which eligible researchers can perform analyses using restricted data resources.*
* ***Through a combination of methods -****Investigators may wish to share their data by a combination of the above methods or in different versions, in order to control the level of access permitted.*

***RCA/DCC guidance:*** ***Data sharing***

*Outline which data you will share and how you will share them, e.g. depositing in a repository, using a secure data service or dealing with data requests individually. The method(s) used will depend upon the size and nature of the data. You should use standards and formats that enable reuse, and ensure data is discoverable through use of accurate metadata and persistent identifiers.*

*The Digital Curation Centre provides useful advice about*[*data appraisal and selection*](http://www.dcc.ac.uk/resources/how-guides/appraise-select-data)*.*

*Most funders allow a delayed release to allow researchers to have exclusive use of their data and to exploit the results of their research. See the RCA page on Research Funder Policies to determine when you need to make your data available. Restrictions on the release of data may be allowed, to protect confidentiality and for other ethical and legal reasons.*

*While restrictions on sharing should be minimised, you should take into account the following when sharing data:*

* *Does your data include confidential and sensitive information?*
* *Have participants given consent for their data to be shared?*
* *Consider what can be done to make sensitive data openly sharable - can these data be anonymised?*
* *Do different parts of your data require different access conditions? These may require separate deposits.*
* *Who will be responsible for controlling access?*

*Whatever form of publishing is used, research data should be licensed to indicate what users may or may not do with the data. Data repositories will indicate what licences are available for the data they house. More information is available from the Digital Curation Centre on* [*how to license research data*](http://www.dcc.ac.uk/resources/how-guides/license-research-data)*.*

*For all Royal College of Art research, a metadata record should be registered in the RCA Research Repository.*

*A Data Access statement should also be included in any publication based upon the research data. A Data Access Statement is a short statement explaining where the data is available, and under what license or access conditions. This helps to further increase the visibility of the data whilst also supporting the validity and reproducibility of your research findings.*

***RCA/DCC guidance: Data repository***

*Long-term preservation and access is generally best managed by using a specialist repository. While you don’t have to specify the repository you will use, you should state the criteria you will use to select it. When considering a repository, you should examine their policies, procedures, metadata standards and any costs that might be incurred. If using a storage facility other than an established repository or data centre, you will need to demonstrate its efficacy and longevity.*

*Some funders specify a data repository, such as*[*UK Data Service ReShare*](http://reshare.ukdataservice.ac.uk/)*,*[*NERC Data Centres*](http://www.nerc.ac.uk/research/sites/data/)*or*[*Archaeology Data Service*](http://archaeologydataservice.ac.uk/)*. Resources such as*[*re3data*](http://www.re3data.org/)*and those provided by* [*BBSRC*](https://bbsrc.ukri.org/research/resources/#datasharing)*or*[*Nature*](https://www.nature.com/sdata/policies/repositories) *can be used to find an appropriate repository. General purpose repositories that you may consider are* [*Zenodo*](https://zenodo.org/) *and* [*Figshare*](https://figshare.com/)*; these are non-discipline specific open access repositories that will ensure the preservation of data for a minimum of 10 years from the last point of access and provide a permanent DOI for the data. Alternatively, RCA researchers can deposit small datasets, particularly those containing textual or vidsual material, in the RCA Research Repository. All research data selected for long-term preservation should be registered in the RCA Research Data Repository, irrespective of where the data files themselves are deposited. Research data in non-digital formats, and digital data that cannot be made accessible or requires controlled access, should also be registered in the RCA Research Repository. This will increase the discoverability and visibility of the research data.*

**4.2 State the timescale for public release of data**

***CRUK Guidance*** *As the value of data is often dependent on its timeliness Cancer Research UK expects that data sharing should occur in a timely manner. Cancer Research UK acknowledges that the investigators who generated the data have a legitimate interest in benefiting from their investment of time and effort and we therefore support the initial investigator having a reasonable period of private use of the data but not prolonged exclusive use.*

*Cancer Research UK expects data to be released no later than the acceptance for publication of the main findings from the final dataset (unless restrictions from third party agreements or IP protection still apply) or on a timescale in line with the procedures of the relevant research area. For example, for crystallography data there is an agreed 12-month delay between publishing the first paper on a structure and making the co-ordinates public.*

*With experiments carried out over an extended period of time, (e.g. population based studies), it is reasonable to expect that subsets of data analysed by the investigator(s) be made available for sharing. The investigator(s) can then continue to benefit from further reasonable periods of exclusive analysis while the dataset as a whole matures.*

***RCA/DCC guidance on Timeframe for Data Sharing***

*Questions to consider:*

* *When will you make the data available?*

*Data (with accompanying metadata) should be shared in a timely fashion. It is generally expected that timely release would be no later than publication of the main findings and should be in-line with established best practice in the field. Researchers have a legitimate interest in benefiting from their investment of time and effort in producing data, but not in prolonged exclusive use. Research funders typically allow embargoes in line with practice in the field, but expect these to be outlined up-front and justified.*

**4.3 Explain any reasons why there may be restrictions on data sharing**

***CRUK Guidance*** *Data which might have the potential to be exploited commercially or otherwise to deliver patient benefit should be discussed with your technology transfer office and Cancer Research Technology prior to data sharing. Cancer Research UK encourages the appropriate filing of patents and recognises that there may be a need to delay the release of data until patent applications have been filed. Whilst there may be a delay in the release of data due to the application process, appropriate intellectual property protection should not hinder data sharing and may be the best way of ensuring that patient (and public) benefit is delivered. Any intellectual property issues or plans for commercialisation that may affect data sharing should be addressed in the data sharing plan. Cancer Research UK understands that unexpected intellectual property may arise during the course of the study and investigators may need to depart from their data sharing plan to protect intellectual property and for any other necessary steps to be taken. Data sharing may also be affected when co-funding is provided by the private sector (e.g. by a pharmaceutical company) or host institution resulting in some restrictions on the disclosure of data. For example with clinical trials, the Trial Management Group and/or trial sponsor etc may impose restrictions on data access. Any restrictions should be outlined in the data sharing plan and applicants should explore ways data sharing requests can be considered by the body that owns the data.*

*e.g. Development arrangements through Cancer Research Technology including intellectual property protection and commercialisation*

*e.g. Proprietary Data - restrictions due to collaborations with for profit organisations International policies governing the sharing of data collected outside of the UK*

*My research seeks supports from both the public and private sectors. How do I deal with the sharing of data? Where research is funded by a commercial sponsor, restrictions on data sharing may apply in arrangements agreed with the sponsor. Any such restriction(s) should be highlighted in the data management and sharing plan. In the event that researchers apply for or receive commercial funding for any part of their research that Cancer Research UK supports they should advise Cancer Research Technology of the situation without delay.*

*e.g. Confidentiality, ethical or consent issues that may arise with the use of data involving human subjects.*

*Investigators carrying out research involving human participants must ensure that consent is obtained to share information; furthermore the necessary legal, ethical and regulatory permissions regarding data sharing should be in place prior to disclosing any data. Every effort must be made to protect the identity of participants and, prior to sharing, data should be anonymised. In addition, any indirect identifiers that may lead to deductive disclosures should be removed to reduce the risk of identification. In most instances, sharing data should be possible without compromising the confidentiality of participants but if there are circumstances where data needs to be restricted due to the inability to protect confidentiality this should be fully addressed in the data management and sharing plan.*

***RCA/DCC guidance: guidance on Restrictions on Sharing***

*Questions to consider:*

* *Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.*
* *What restrictions are needed and why?*
* *What action will you take to overcome or minimise restrictions?*

*Outline any expected difficulties in data sharing, along with causes and possible measures to overcome these. Restrictions to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include: anonymising or aggregating data; gaining participant consent for data sharing; gaining copyright permissions; and agreeing a limited embargo period.*

***RCA/DCC guidance on IPR Ownership and Licensing***

*State who will own the copyright and IPR of any existing data as well as new data that you will generate. For multi-partner projects, IPR ownership should be covered in the consortium agreement.*

*Outline any restrictions needed on data sharing, e.g., to protect proprietary or patentable data.*

*Explain how the data will be licensed for reuse. See the DCC guide on*[*How to license research data*](http://www.dcc.ac.uk/resources/how-guides/license-research-data)*and EUDAT’s*[*data and software licensing wizard*](https://ufal.github.io/public-license-selector/)*.*

### 7. Preservation Plan

**State the long-term preservation plan for the dataset**

***CRUK Guidance*** *Once the funding for a project has ceased researchers should preserve all data resulting from that grant to ensure that data can be used for follow-up or new studies. Cancer Research UK expects that data be preserved and available for sharing with the science community for a minimum period of five years following the end of a research grant.*

***RCA/DCC guidance:*** ***Preservation***

*Describe how you will preserve and share your data, including the length of time they will be kept and the nature of the storage location. The RCA Research Data Management Policy requires that all data needed to validate research findings are kept for a minimum of 10 years. Also indicate if any additional resources or funding will be required to deposit and store the data.*

*Research funders generally expect data with long-term value to be preserved and remain accessible, alongside the software and code needed to reproduce your findings. This does not mean that you need to keep all of your data, but you will need to state who will be responsible for choosing and archiving data, as well as documenting the removal of any data that must be destroyed.*

*It is particularly important to preserve data which cannot be remeasured or recreated. Many research funders specify which data need to be preserved, how long for and where they should be deposited. See the DCC guide*[*How to appraise and select research data for curation*](http://www.dcc.ac.uk/resources/how-guides/appraise-select-data)*.*