CRUK Population Studies DMP Assessment Rubric v2.0

Funder Template	CRUK - Population research.			
Purpose of rubric	Providing feedback to researchers			
Notes	CRUK have subject-specific guidance in the following areas: Discovery research (http://www.cancerresearchuk.org/sites/default/files/hands_on_data_sharing_advice_-_basic_science.pdf) Clinical research (http://www.cancerresearchuk.org/sites/default/files/hands_on_data_sharing_advice_-_clinical.pdf) Population research (http://www.cancerresearchuk.org/sites/default/files/hands_on_data_sharing_advice_-_population.pdf)			
Documents Used	Cancer Research UK Policy on Data Sharing and Preservation Cancer Research UK Practical guidance for researchers on writing data sharing plans (http://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/practical-guidance-for-researchers-on-writing-data-sharing-plans) CRUK Template for a Data Management and Sharing Plan (Population research) - Spetember 2014			
Version history	v0 - basic document in development v1.0 - version circulated to funder for comment v1.1 - revision in response to funder comments v2.0 - formatting edits for download.			
		Performance Levels		
		Detailed	Addressed but incomplete / unsatisfactory	Not addressed
Section 1.1	Type of Study?	Plan concisely summarises the type of study.	Not applicable	Type of study is not mentioned.
Section 1.2	What types of data will be managed?	Data types clearly defined. Eg imaging data, genotypic data, clinical measurements, survey data, interviews, medical records, tissue samples, qualitative data / quantitative data etc. It is important to clearly state which data can be shared, which data cannot be shared and why.	Data types mentioned for some of project / dataset but not all. No indication as to which data may or may not be shared. Reasons for data sharing suitability might be missing.	Data types are not mentioned.
Section 1.3	What scale / volume of data will be managed?	Clear estimate of dataset size and number of records is given for each data type.	Dataset size given but not broken down by data type. Size not give for all data types. Dataset size is clearly unrealistic (not always possible to judge!).	Dataset size is not mentioned.
Section 1.3	What format of data will be managed?	Data formats and software used are clearly defined. Eg spreadsheets in .csv or .xlsx; micrographs in .tiff or .jgp; proprietary manufacturer formats where necessary.	Data formats are mentioned for some of dataset but not all.	Dataset formats are not mentioned.
Section 1.3	Do formats enable sharing and long-term validity of data?	Clear assessment of whether file formats are widely used in the field and will be usable over the long-term.	Some assessment of whether file formats are widely used in the field and will be usable over the long term, but information is missing for some data types or formats, or is vague.	Reasons for selecting / using formats is not mentioned.
Section 2.1	How will data be stored?	Clear description of data storage systems. Eg departmental server, on machine, on portable hard-drive. Ideally this section should include an assessment of the suitability of the storage and the security implications for the data.	Mention of data storage systems, but lacking detail or clearly inappropriate (could be difficult to judge).	Dataset storage is not mentioned.
Section 2.1	How will data be backed up?	Clear description of data backup routines / protocols. Eg automatic backup every night; weekly backup of equipment to server.	Some mention of data backup routines / protocols but detail lacking or clearly inappropriate.	No mention of data backup systems.
Section 2.1	How will data be curated/managed during the project?	Clear data management processes are either described or referenced.	Data management processes are mentioned but are not clear or do not cover all parts of dataset.	Data management processes are not mentioned.
Section 2.1	Which formal or community standards for data will be used?	Data management methodology is clearly stated or referenced. eg file naming conventions, file architecture etc.	Methodology is mentioned for a subset of the data to be collected	No methodology is mentioned.

Section 2.2	How will the dataset be documented?	Clear outline of documentation which will be maintained. Eg methods documentation, analytical and procedural information, provenance of data and their coding, detailed descriptions for variables, records etc.	Some mention of documentation but lacking detail or doesn't cover all of dataset.	Accompanying documentation is not mentioned.
Section 2.2	What metadata will be supplied with dataset?	Clear outline of metadata strategy with references to existing good practice in the community eg MIAME for microarray, CDISC for clinical research or capturing instrument metadata alongside data. Detailed project-specific approach where community standards don't exist.	Some mention of metadata without detail about community standards or a project-specific approach.	Metadata is not mentioned.
Section 2.3	How long with the dataset be retained for?	Clear data retention statement.	Clear data retention statement but does not cover all parts of dataset or is at odds with funder requirements.	No mention of period for data retention.
Section 2.3	Which data will be retained?	Clear explanation of selection and appraisal strategy, or list of sub-datasets which will be retained. Clear indication of which data may not be retained.	Explanation of selection and appraisal strategy, or list of sub-datasets, but doesn't cover full dataset or is clearly inappropriate (could be difficult to judge).	No mention of selection and appraisal or dataset retention.
Section 2.3	What is the long-term preservation strategy for the dataset?	Clear strategy for long-term preservation of dataset. Eg deposit in an appropriate responsible repository. Clear statement that dataset won't be preserved / is not suitable for preservation.	Preservation is mentioned but strategy is not clear or lacks detail.	No mention of preservation of dataset.
Section 2.3	Which formal preservation standards will be used?	Clear reference to preservation standard, eg ISO 14721:2012. Or an explanation as to why a formal preservation standard is not appropriate.	An indication that preservation standard(s) will be used, but no specification as to which one.	No mention of formal preservation standards.
Section 3	What are the main risks to the confidentiality and security of information related to human participants?	Thorough assessment of potential for breaches in data security. Eg theft of IT equipment, unauthorised access to or hacking of server / cloud storage, loss or theft of data storage devices (drives of disks), accidental release of identifying information	Some consideration of potential for breaches in data security, but not realistic or comprehensive.	No consideration of potential for breaches in data security.
Section 3	How will these risks be managed?	For each identified risk to data security, there should be an associated strategy for managing the risk. Eg ensuring that offices are locked when not in use; password protection on all storage drives; encryption of data when on portable storage or when being transferred between study staff; double-checking of communications to ensure identifying data is not released; not using email attachments to send sensitive data; not storing data on drives outwith institutional control.	General strategies to protect personal and sensitive data are outlined, but detail is missing or strategies do not cover all the risks identified.	No consideration given to how to manage risks to personal and sensitive data.
Section 3	Which formal data security standards will be used?	Clear reference to data security standard, eg ISO27001. Or an explanation as to why a formal data security standard is not appropriate.	An indication that a security standard will be used, but no specification as to which one. Details of security standards for some subsections of the dataset, but does not cover all data.	No mention of formal data security standards.
Section 4.1	Which data repositories will be used for data deposit?	For each dataset, a responsible repository will be identified which is suitable for that dataset. Consideration will be given to ingest criteria for each repository and how IP and proprietary data are handled.	Repositories are identified for some subsets of the data to be generated, but some subsets of data are unaccounted for. Or, data deposit in a responsible repository is mentioned, but repository is not identified.	No mention of suitable repositories for any of the data produced.

Section 4.2	How will potential new users find out about the dataset?	Clear strategy for publishing information on dataset. This could include data citation in paper, indexing in searchable data registries (eg Datacite), inclusion of dataset info on personal / institutional / project webpages. Or, a clear statement that the data generated won't be suitable for reuse, so potential new users don't need to be considered.	Mention that a strategy for publicising the dataset exists or will be developed but no detail. Strategy only covers some of the datasets to be generated.	No mention of how potential new users will be able to find information about the dataset(s).
Section 4.2	Will the data sharing position be published on study website?	Clear statement indicating that rules governing access to research data will be published on study website. If users cannot access information about the data sharing position on the study website, a reason should be given and an alternative mechanism described.	Mention of data sharing position, but no clear indication as to where this can be accessed.	No mention of how potential new users will be able to find out about the data sharing position of the study.
Section 4.3	What is the policy on exclusive use of the data?	A clearly defined period for exclusive use is given along with a justification for that period. This may depend upon the nature and value of the data and the way in which they are generated and used. Alternatively, there is a clear statement that a period for exclusive use is not required (this is unlikely, but possible).	A period for exclusive use is mentioned, but is not clearly defined, or is lacking justification.	A period for exclusive use is not mentioned.
Section 4.4	What are the potential restrictions to data sharing?	There is a clear assessment of any ethical, IPR or patient confidentiality concerns. Alternatively, there is a clear statement that there are no restrictions on this dataset.	Data sharing restrictions or problems are mentioned, but detail is lacking.	Data sharing restrictions are not mentioned.
Section 4.4	What plans are in place to limit potential restrictions to data sharing?	Strategies to limit data sharing restrictions are discussed. Eg anonymisation or aggregation of data; participant consent for data sharing; gaining copyright permissions; material transfer agreements (MTAs); restricted or controlled access to data.	Strategies to limit data sharing restrictions are mentioned, but do not cover all the restrictions identified or are unsuitable or impractical (this could be difficult to assess).	No strategies to limit data sharing restrictions are discussed.
Section 4.5	What are the milestones for sharing?	Schedule of milestones is explained clearly, and may include expected publication times.	Milestones are mentioned but schedule isn't clear or periods of time may not be accounted for.	Milestones are not mentioned.
Section 4.6	Who will make decisions on supplying data to new users?	An individual or committee within the study is identified, either by name or position. Contact details are included. Alternatively, it should be indicated that as the data will be shared via a third party repository.	An individual or committee within the study is identified, but contact details are not given. It is indicated that there will be someone overseeing data access, but no individual or committee is identified. There is no indication that the data will be shared by a third party repository.	No indication is given that responsibility for overseeing data access will be assigned to an individual or committee.
Section 4.6	How will independent oversight of data access and sharing work?	The process by which a potential user applies for access to data, and the criteria that the application needs to meet should be outlined. The criteria that must be met for access to shared data should match the data sharing position published on the study website.	The process by which a user applies for access to shared data is mentioned, but detail is missing or there is a mismatch between the process outlined in this section and the position on data sharing that is outlined on the study website.	The process for independent oversight of data sharing and access is not mentioned.
Section 4.7	Will data sharing agreements be necessary?	Datasets to which access would require a data sharing agreement should be identified. Purposes for which data sharing agreements are required should be identified. Alternatively, a statement that data sharing agreements are not needed should be included.	It is indicated that data sharing agreements might be needed, but the relevant datasets or purposes which would require this are not identified.	No indication is given as to whether data sharing agreements would be needed or not.
Section 4.7	What would be main responsibilities of external data users?	If data sharing agreements are indicated as necessary, the main responsibilities of the external data user should be clearly listed.	The responsibilities of external data users are mentioned but detail is missing. Alternatively, responsibility of external data users are only listed for a subset of the data that requires data sharing agreements.	The responsibilities of external data users are not mentioned.

Section 5	Who is responsible for study-wide data management?	Clear indication of who is responsible. This might be more than one person.	Mentions that responsibility will be taken for data management without giving details of who.	No mention of responsibility for data management.
Section 5	Who is responsible for specific data management tasks?	Clear indication of who has responsibility for particular data management tasks. Eg record keeping, data entry, metadata recording, experimental protocols. Or, might indicate that the person responsible for overall data management will also take responsibility for all specific tasks (this is unlikely, and unlikely to be desirable, in large studies).	Mentions that responsibility will be taken for data management without giving details of who is responsible for which processes.	No mention of responsibility for particular data management processes.
Section 6	What are the relevant institutional, departmental or study policies?	A list of relevant policies is provided.	A partial list of relevant policies is provided.	No relevant policies are mentioned.