**Proposed revised set of themes for Data Management Plans**

A revised set of themes for Data Management Plans, put out for community consultation in September 2016. For more details about the themes and how the DCC and UC3 plan to use them to support our objectives around machine-actionable DMPs, please see the associated blog post: http://www.dcc.ac.uk/blog/common-set-themes-dmps-seeking-input

| Theme | DCC and UC3 guidance |
|---|---|
| DATA DESCRIPTION | **Questions to consider:**<br>- What types of data will you collect or create?<br>- Are there any existing data or methods that you can reuse?<br>- Are there any restrictions on the reuse (and subsequent sharing) of third-party data?<br><br>**Guidance:**<br>Provide a summary of the data that will be created, noting the content, coverage and data type e.g. tabular data, survey data, experimental measurements, models, software, audiovisual data, physical samples, etc.<br>Consider how your data could complement and integrate with existing data, or whether there are any existing data that you could reuse. If purchasing or reusing existing data sources, explain how issues such as copyright and IPR have been addressed. |
| DATA FORMAT | **Questions to consider:**<br>- What format(s) will your data be in?<br>- Why have you chosen to use particular formats?<br>- Do the chosen formats and software enable data sharing?<br><br>**Guidance:**<br>Outline and justify your choice of format e.g. plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw). Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. Using standardised, interchangeable or open formats ensures the long-term usability of data.<br>See UK Data Service guidance on recommended formats or DataONE Best Practices for file formats |
| DATA VOLUME | **Questions to consider:**<br>- What volume of data will you create?<br>- Do you have sufficient storage or should you include costs for more?<br>- Will the scale of the data pose challenges when sharing or transferring data between sites?<br>- Have you consulted with a data repository to determine preservation costs?<br><br>**Guidance:**<br>Consider the implications of data volumes in terms of storage, backup and access. Estimate the volume of data in MB/GB/TB and how this will grow to make sure any additional storage and technical support required can be met. |
| DATA QUALITY | **Questions to consider:**<br>- What standards or methods will you use to create the data? |

| | |
|---|---|
| | - What quality assurance processes will you adopt?<br>- How will you organise, version and structure the data?<br>- How will the data be processed?<br><br>**Guidance:**<br>Outline how the data will be collected/generated and which community data standards (if any) will be used. 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. Indicate how the data will be organised during the project, mentioning for example naming conventions, version control and folder structures. Consistent, well-ordered research data will be easier to find, understand and reuse.<br>See the DataOne Best Practices for data quality |
| METADATA | **Questions to consider:**<br>- What metadata, documentation or other information should accompany the data?<br>- How will you capture / create the metadata and documentation?<br>- What metadata standards will you use and why?<br>- What file-naming conventions will you use?<br><br>**Guidance:**<br>Metadata should be created to describe the data and aid discovery. 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. Researchers are strongly encouraged to use community metadata standards where these are in place. The Research Data Alliance offers a Directory of Metadata Standards.<br>Also describe the types of documentation that will accompany the data to provide secondary users with any necessary details to prevent misuse, misinterpretation or confusion. 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, software used to collect and/or process the data. |
| ETHICS & PRIVACY | **Questions to consider:**<br>- Have you gained consent for data sharing and preservation?<br>- How will sensitive data be handled to ensure it is stored and transferred securely?<br>- How will you protect the identity of participants? e.g. via anonymisation or using managed access procedures<br><br>**Guidance:**<br>Investigators carrying out research involving human participants must ensure that consent is obtained to share data. Managing ethical concerns may include anonymisation of data, referral to departmental or institutional ethics committees, and formal consent agreements. Ethical issues may affect how you store data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.<br>See UK Data Service guidance on consent for data sharing<br>ICPSR approach to confidentiality<br>Health Insurance Portability and Accountability Act (HIPAA) regulations for health research |
| INTELLECTUAL PROPERTY RIGHTS | **Questions to consider:**<br>- Who owns the data?<br>- How will the data be licensed for reuse?<br>- Will data sharing be postponed / restricted e.g. to seek patents? |

| | |
|---|---|
| | **Guidance:**<br>State who will own the copyright and IPR of any 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.<br>See the DCC guide on How to license research data and EUDAT's data and software licensing wizard. |
| STORAGE & BACKUP | **Questions to consider:**<br>- Where will the data be stored during the course of research activities?<br>- How will the data be backed up? i.e. how often, to where, how many copies, is this automated.<br>- Who will be responsible for storage and backup?<br>- Do you have access to enough storage or will you need to budget for additional services?<br><br>**Guidance:**<br>Describe how the data will be stored and backed up to ensure the data and metadata are secure during the lifetime of the project. Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage with automatic backup, for example that provided by university IT teams, is preferable.<br>See UK Data Service Guidance on data storage or DataONE Best Practices for storage |
| DATA SECURITY | **Questions to consider:**<br>- What are the risks to data security and how will these be managed?<br>- Will you follow any formal standards?<br><br>**Guidance:**<br>If your data is sensitive (e.g. detailed personal data, politically sensitive information or trade secrets) you should discuss appropriate security measures. Note the main risks to security and how these will be managed. Identify any formal standards that you will comply with e.g. ISO 27001.<br>See the DCC Briefing Paper on Information Security Management - ISO 27000 and UK Data Service guidance on data security. |
| DATA SHARING | **Questions to consider:**<br>- Which data will you share, and under what conditions?<br>- How will you make the data available to others?<br>- Who may be interested in using your data?<br>- How will potential users find out about your data?<br><br>**Guidance:**<br>Consider where, how, and to whom the data should be made available. Will you share data via a data repository, handle data requests directly or use another mechanism? The methods used to share data will depend on a number of factors such as the type, size, complexity and sensitivity of data. Consider the possibilities for reuse of your data in other contexts. Where there is potential for reuse, you should use standards and formats that facilitate this, and ensure that appropriate metadata is available online so your data can be discovered. You should always review any relevant funder requirements. |
| TIMEFRAME FOR RELEASE | **Questions to consider:**<br>- When will you make the data available?<br>- Will you need exclusive use of the data for a limited period? |

| | |
|---|---|
| | - If you need an embargo, please explain why.<br><br>**Guidance:**<br>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. Research funders typically allow embargoes in line with established best practice in the field, but not prolonged exclusive use. You should always review any relevant funder requirements. |
| RESTRICTED-USE DATA | **Questions to consider:**<br>- Are any restrictions on data sharing required? e.g. limits on who can use the data, when and for what purpose.<br>- How will this data be shared? e.g. by using data enclaves, secure data services, data sharing agreements<br>- What actions will you take to overcome or minimise restrictions?<br><br>**Guidance:**<br>Outline any restrictions on data sharing. These may be due to participant confidentiality, consent agreements or IPR. Indicate whether external users will be bound by data sharing agreements, licenses or end-user agreements. If so, set out the terms and key responsibilities to be followed. Note how access will be controlled, for example by the use of specialist services. A data enclave provides a controlled secure environment in which eligible researchers can perform analyses using restricted data resources.  Also consider potential strategies to overcome or minimise restrictions on sharing. These may include anonymising or aggregating data, gaining participant consent for data sharing, gaining copyright permissions, and agreeing a limited embargo period. |
| DATA REPOSITORY | **Questions to consider:**<br>- In which repository will the data be deposited?<br>- Have you consulted with the repository to understand their policies and procedures?<br><br>**Guidance:**<br>Most research funders recommend the use of established data repositories, community databases and related initiatives to aid data preservation, sharing and reuse. An international list of data repositories is available via Re3data and some universities or publishers provide lists of recommendations e.g. PLOS ONE recommeded repsositories. |
| PRESERVATION | **Questions to consider:**<br>- Which data are of long-term value and should be shared and/or preserved?<br>- How long will the data be retained?<br>- What is the long-term preservation plan for the data? e.g. deposit in a data repository<br>- Will additional resources be needed to prepare data for deposit or meet any charges from data repositories?<br><br>**Guidance:**<br>Indicate which data you intend to preserve and where relevant, state for how long. Outline the plans for preparing the data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that the data can be curated effectively beyond the lifetime of the grant.<br>See the DCC guide: How to appraise and select research data for curation or DataONE Best Practices: Identifying data with long-term value |
| ROLES & RESPONSIBILITIES | **Questions to consider:**<br>- Who is responsible for each data management activity?<br>- How are responsibilities split across partner sites in collaborative research projects? |

| | |
|---|---|
| | **Guidance:**<br>Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Individuals should be named where possible. For collaborative projects you should explain the co-ordination of data management responsibilities across partners.<br>See UK Data Service guidance on data management roles and responsibilities or DataONE Best Practices: Define roles and assign responsibilities for data management. |
| BUDGET | **Questions to consider:**<br>- What additional resources are needed to deliver your plan?<br>- Is specialist expertise (or training for existing staff) required?<br>- Do you have sufficient storage and equipment or do you need to budget for more?<br>- Will charges be applied by data repositories?<br>- Have you included costs for time and effort to prepare the data for sharing / preservation?<br><br>**Guidance:**<br>Carefully consider any resources needed to deliver the plan and include a justification for any additional costs. Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. Provide details and justification for any hardware or software which will be purchased or additional storage and backup costs that may be charged by IT services. Funding should be included to cover any charges applied by data repositories, for example to handle large volumes of data. Also remember to include costs for time and effort to prepare data for deposit and ensure it is adequately documented to enable reuse. If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data.<br>See UK Data Service guidance on costing data management |
| RELATED POLICIES | **Questions to consider:**<br>- Are there any existing procedures that you will base your approach on?<br>- Does your department/group/institution have relevant data policies or guidelines?<br>- Are there any formal standards that you will adopt?<br><br>**Guidance:**<br>List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP may be determined by the content of other policies. If so, point/link to them here. |