Data Management Planning (DMP) themes

This is a consolidated set of Data Management Planning themes and guidance, issued by the Digital Curation Centre (DCC) and the University of California Curation Centre (UC3) in December 2016. The revisions are based on community feedback and helped us to consolidate an original, longer set of themes used in DMPonline. Details on the consultation process and future plans are available in the associated blog post: http://www.dcc.ac.uk/blog/dmp-themes-and-then-there-were-14

Theme	DCC & UC3 guidance
DATA DESCRIPTION	• Give a summary of the data you will collect or create, noting the content, coverage 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.
DATA FORMAT	• Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tfw). • Explain why you have chosen certain formats. 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; these are recommended for sharing and archiving. • See UK Data Service guidance on recommended formats or DataONE Best Practices for file formats
DATA VOLUME	• Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports). • Consider the implications of data volumes in terms of storage, access and preservation. Do you need to

	include additional costs? • Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?
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. • See the DataOne Best Practices for data quality
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. 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.
ETHICS & PRIVACY	• Investigators carrying out research involving human participants should request consent to preserve and share the data. Do not just ask for permission to use the data in your study or make unnecessary promises to delete it at the end. • Consider how you will protect the identity of participants, e.g., via anonymisation or using managed access procedures.

	• Ethical issues may affect how you store and transfer 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. • See UK Data Service guidance on consent for data sharing • See ICPSR approach to confidentiality and Health Insurance Portability and Accountability Act (HIPAA) regulations for health research
INTELLECTUAL PROPERTY RIGHTS	• 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 and EUDAT's data and software licensing wizard.
STORAGE & SECURITY	• Describe where the data will be stored and backed up during the course of research activities. This may vary if you are doing fieldwork or working across multiple sites so explain each procedure. • Identify who will be responsible for backup and how often this will be performed. The use of robust, managed storage with automatic backup, for example, that provided by university IT teams, is preferable. Storing data on laptops, computer hard drives or external storage devices alone is very risky. • See UK Data Service Guidance on data storage or DataONE Best Practices for storage • Also consider data security, particularly if your data is sensitive e.g., detailed personal data, politically sensitive information or trade secrets. Note the main risks and how these will be managed. Also note whether any institutional data security policies are in place. • Identify any formal standards that you will comply with, e.g., ISO 27001. See the DCC Briefing Paper on Information Security Management - ISO 27000 and UK Data Service guidance on data security
DATA SHARING	• How will you share the data e.g. deposit in a data repository, use a secure data service, handle data requests directly or use another mechanism? The methods used will depend on a number of factors such as the type, size, complexity and sensitivity of the data. • When will you make the data available? Research funders expect timely release. They typically allow embargoes but not prolonged exclusive use.

	• Who will be able to use your data? If you need to restrict access to certain communities or apply data sharing agreements, explain why. • Consider strategies to 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. • How might your data be reused 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. Persistent identifiers should be applied so people can reliably and efficiently find your data. They also help you to track citations and reuse.
DATA REPOSITORY	• Where will the data be deposited? 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. • It helps to show that you have consulted with the repository to understand their policies and procedures, including any metadata standards, and costs involved. • An international list of data repositories is available via Re3data and some universities or publishers provide lists of recommendations e.g. PLOS ONE recommended repositories
PRESERVATION	• Outline the plans for data sharing and preservation - how long will the data be retained and where will it be archived? • Will additional resources be needed to prepare data for deposit or meet any charges from data repositories? 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	• 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 coordination of data management responsibilities across partners. • See UK Data Service guidance on data management roles and responsibilities or DataONE Best Practices:

	Define roles and assign responsibilities for data management
BUDGET	• Carefully consider and justify any resources needed to deliver the plan. These may include storage costs, hardware, staff time, costs of preparing data for deposit and repository charges. • Outline any relevant technical expertise, support and training that is likely to be required and how it will be acquired. • If you are not depositing in a data repository, ensure you have appropriate resources and systems in place to share and preserve the data. See UK Data Service guidance on costing data management
RELATED POLICIES	• Consider whether there are any existing procedures that you can base your approach on. If your group/department has local guidelines that you work to, point to them here. • List any other relevant funder, institutional, departmental or group policies on data management, data sharing and data security.