|  |  |
| --- | --- |
| 1. Overview |  |
| 1. **Researcher’s Name (Principal investigator):** | Research Data Management Plan  (v6) |
| 1. **Title and ID of Research Project:** |
| 1. **Length of Project**   **Start Date:**  **End Date:** |
| 1. **A brief statement of the aim(s) of the project** 2. **Funding body and programme:** 3. **Related policies:** 4. **Version of DMP:** |

|  |  |
| --- | --- |
| 1. Data collection and documentation | **Help, Guidanceand *Example Answers*** |
| 1. **Briefly describe the data that you will collect for the project** | * Details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media. Details on existing datasets to be re-use. Methods by which data will be collected or created   I record interviews with subjects using a digital audio recorder, then transcribe them into text. I test my catalyst under a number of conditions, then submit samples of the products to analysis facilities. I generate data using model code that I’ve written, then process it various ways to produce visualisations. I combine existing data from a number of sources [e.g.…] and reanalyse them to derive new conclusions. |
| 1. **Frequency of new data (how often will you get new data and over what time period?)** | All of my data will come from a single 3-month field trip in my second year. I expect to run two or three experiments each week through my second year and much of my third year – about 100 in total. |
| 1. **Quantity of data (Terabytes, other forms of storage)** | * Details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).   Each experiment produces about 50MB of data, so over the course of my PhD I expect this to add up to about 5GB. 2 drawers of a standard filing cabinet |
| 1. **What format is the data in?** | * Details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf). * Explain why certain formats have been chosen and indicate if they are in open and standard format. Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such ‘preferred formats’).If a proprietary format is used, it explains why. * Clearly state, if applicable, that no new data will be produced or generated by the project.   *Binary data.*  *The format used by a particular instrument paper notebook.*  *Audio recordings from interviews will be stored in MP3 format.*  *Survey data will be stored in SPSS format.* |
| 1. **Describe the system to name and structure any electronic files and records.** | * UK Data Service guidance on organising files: [*https://www.ukdataservice.ac.uk/manage-data/format/organising.aspx*](https://www.ukdataservice.ac.uk/manage-data/format/organising.aspx)   I use the structure <thesis chapter>/<date>-<experiment number>. A folder for each project phase, and within those a folder for each interview. Each filename starts with the date on which the data was collected  Interview transcripts will be stored as separate files and all 20 files will be placed in a single folder. |
| 1. **Could the data be considered high value and/or vulnerable? E.g. is your data likely to attract “hactivists”? How could this be mitigated? What measures will you take to comply with the security requirements and to mitigate the risks?** | * Describe how you can restore your data in the event of data loss and who is responsible. If applicable, please describe procedures to ensure personal data are handled confidentially and who is responsible. Precaution examples: Access restrictions (physical or digital), Encryption, Reduce data sensitivity, Regular and timely back-ups, Master (locked) copy stored on university network storage, Master (locked) copy stored elsewhere, Data handling procedures and/or training for data handlers, name any other.   My data is of high value and may be subject to threats from overseas governments, I will check for advice in the *Information Security  Awareness training Canvas course* <https://birmingham.instructure.com/courses/5162> |
| 1. **What standard will you use to describe your data?** | * Please refer to any metadata standards in your field if they exist. And Where will metadata be registered?   *I have a discipline-specific metadata standard, namely:…* |
| 1. **What data quality control measures will be used?** | * Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies*.*   *The metadata will be recorded to a separate README file*  *I will describe in a txt file what is in a file, where did it come from, how could it be retrieved if needed, any existing problems etc*. |
| 1. **What additional information is necessary to understand the data? E.g. abbreviations, supplementary notes.** | I keep additional notes about interviews and participants in a Word document with the audio recordings and transcripts. Abbreviations used for column headings are kept in a separate text document. The content of digital photographs are recorded in the file name.  Database schema  Laboratory notebooks  Software syntax and output files |

|  |  |
| --- | --- |
| 1. Data Storage and Backup | Help, Guidance and *Example Answers* |
| 1. **What different versions of the data do you create? E.g. versions of data files** | * Version your files e.g. by using a 'revision' numbering system. Any major changes to a file can be indicated by numbers, for example, v01 would be the first version, v02 the second version * Consider using version control software such as Git which can be used on text files as well as code. There is a Library Carpentry session on Introduction to Git: <https://librarycarpentry.org/lc-git/> * More details in Version control and authenticity from the UK Data Service.   As I survey new cohorts, data is appended to the dataset and saved as a new file. There is only ever one version of each data file — new experiments create new data, which is stored in a new set of files. Each time I run a new version of my model, intermediate files are written over, but the final results are saved as a new file |
| 1. **Where will the data be stored?** | * Describe where the data will be backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations. * Electronic data should be stored in multiple geographically distinct locations and any sensitive data on local / removable drives should be encrypted. Note that the University Research Data Store is backed up to two physically separate data centres.   My primary copy is on the university X: drive, and I copy files to my laptop to work on while away from the office. |
| 1. **Describe the regime for backing up the data.** | * How will data and metadata be stored and backed up during the research? Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended*.*   Working data is backed up on the UoB Research Data Store. I make sure I copy the latest versions of my working files there each day. I regularly scan my paper-notebook and store digital copies on the X:drive |
| 1. **Describe the procedure to be used to ensure files can be restored from the backups.** | * Backing up data is an essential practice to insure against the loss of valuable information. Checke the information at the UoB [Backup & Retention Policy](https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/research-data-service/RDS/BackupRetentionPolicy.aspx)   Weekly check that files on the X: drive are still usable.*The data will be stored on University of Birmingham (UoB) storage (BEAR Research Data Store), accessed via network shares mounted on the computers of the Principal Investigator (PI), Co-PI’s and other researchers. Backups are made overnight from the Research Data Store (RDS) and any files that are created or changed that day will be backed up. Backups are also copied to a second location for disaster recovery purposes.* |
| 1. **Are there any non-digital data or outputs that the project will generate? Where will these outputs be stored?** | * Do you have a protocol for the storage and deletion of non-digital data? Please specify briefly and describe who is responsible for storage of these outputs.   Signed consent forms will be stored in a locked cabinet in the office. |
| 1. **How will data security and protection of sensitive data be taken care of during the research?** | * Which institutional and/or national data protection policies are in place and provides a link to where they can be accessed who will have access to the data during the researchand how access to data is controlled, especially in collaborative partnerships. * A Data Protection Impact Assessment (DPIA) must be carried out by the relevant School or Professional Service and approved by the Data Protection Officer (or nominee). This must be retained and regularly reviewed and revised as necessary for the duration of the processing. For more information follow the link here: https://www.birmingham.ac.uk/documents/university/legal/data-prot-policy.pdf * Clearly describe the additional security measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe, when sensitive data are involved (for example, personal data, politically sensitive information, or trade secrets). * BEAR Services provide [The Data Matrix](https://kb.bham.ac.uk/KB15064) a great guidance on sensitive data |
| 1. **What are the main risks to data security and what would happen if the data got lost or become unusable?** | * Examples of risks: Accidental deletion or file corruption, Theft of, or damage to, equipment, Overwriting or version loss Data leak, unauthorised access, or unauthorised use, or explain any other   *“Original versions of files will always be kept on the server. If copies of files are held on a laptop and edits made, their file name will be changed”.* |

|  |  |
| --- | --- |
| 1. Ethics and legal compliance | Help, Guidance and *Example Answers* |
| 1. **Who owns the data? (Copyright and intellectual property rights)** | * Explain who will be the owner of the data, meaning who will have the rights to control access: * Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement. For clarification on who owns the data, consult the University’s [Code of Practice for Research](https://www.birmingham.ac.uk/Documents/university/legal/research.pdf), and/or discuss with your Principal Investigator. Also check your agreements or contracts with any sponsors or collaborators.   I own the copyright of the newly created research data.  The media information is copyright of the respective publishers and will be attributed accordingly |
| 1. **Are there restrictions on who can use the data, and if so, what are they?** | * Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses. More information: <https://intranet.birmingham.ac.uk/as/libraryservices/library/copyright/licences.aspx>Clearly explain, if applicable, why data sharing is limited or not possible, and who can access the data under which conditions (for example, only members of certain communities or via a sharing agreement). * Describe the procedure to manage access to only authorised users.   All my data is covered by a confidentiality agreement and cannot be shared. Some of my data identifies individual patients and must be anonymised before sharing.  "Personally identifying information (PII) will be removed and stored separately from the data files. Access to these separately stored PII files will have the added protection of encryption, such as via a password." |
| 1. **If personal sensitive data are processed, how will compliance with legislation on personal data and security be ensured?**   . | * Ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with: * Gain informed consent for preservation and/or sharing of personal data. * Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).Check the following tool: <https://www.openaire.eu/item/amnesia-data-anonymization-made-easy> * And the guidance from UoB IT services: <https://intranet.birmingham.ac.uk/it/teams/infrastructure/research/bear/BEARDataShare/sensitive-data.aspx>   ”To ensure anonymity all personal identifiers will be removed from the dataset before sharing”. |

|  |  |
| --- | --- |
| 1. Archiving (preserving the data for future use) | Help, Guidance and *Example Answers* |
| 1. **What data should be kept beyond the end of the project and under what requirements (contractual, legal or regulatory)?** | * Explain the foreseeable research uses (and/ or users) for the data.   All data, both raw and processed. Only simulation code and input parameters. Transcripts of all interviews, but not recordings.  *Data underlying publications* |
| 1. **How long should it be kept?** | * This should be in alignment with funder, institutional, or national policies and/or legislation, or community standards. * Data will be preserved in the long term and clearly indicate for how long.   Until the end of my PhD For 10 years after the end of the project |
| 1. **Where will the data be archived?** | * Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.   In the UK Data Archive On the Ubira eData repository. |
| 1. **Who will create and maintain the archive of data?** | * Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?   I am responsible for archiving data, and the archive service will maintain it. My supervisor will deal with this. |
| 1. **Are there restrictions on who can access the archived data?** | Data will be embargoed for 12 months to enable patent protection. Full data will never be public as it contains sensitive personal information, but anonymised data will be made available on request. |
| 1. **If the data can be made openly available, at what point can this happen?**   **How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?** | * Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism). Include any licence under which the data will be shared (e.g. CC-BY for maximum capacity for re-use) * Outline the plan for data preservation and give information on how long the data will be retained. * Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents. |
| **g. How will you enable findability and availability for the long term and maximise data’s discoverability?** | * Identify where the data will be deposited, how the metadata will be made available and whether it will be given a persistent identifier such as a DOI.   For the sharing of data with the general public, experimental data that underpins published work will be deposited in trusted digital repositories.  UoB uses PURE, as it’s research information and management system; a metadata record will be created in PURE for the data set. PURE exposes metadata to web services via standard metadata harvesting protocols.  *My data will be deposited in the UBIRA edata repository (based on Eprints) and be given a DOI to aid long term discoverability*  *I will deposit data in a trusted data repository (e.g. DANS Easy, 4TU.ResearchData) as indicated below:*  *According to the data protocol of my institute, I will archive data in the data repository indicated below (e.g.Ubira eData):*  *I will deposit data in a discipline-specific data repository as indicated below:*  *I will use an archive specifically for my collaboration, namely:*  *I will not use a data repository and will explain below how I will make my data findable and accessible for the long term.*  *I will not make my data findable and accessible and I will explain why* |
| 1. **What are the likely (estimated) costs of preserving the data?** | * **What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?** * Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. * Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges. * Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered. * Consider the likely amount of storage you will need, and how long this will need to be archived for. The University of Birmingham provides the BEAR Archive for storing valuable data associated with your research project which may be needed in the future but which is not currently being actively used and is not needed to support a project. Up to 20TB of data storage is provided free of charge for up to 10 years. For more information see https://intranet.birmingham.ac.uk/bear-archive   *I will make use of the FAIR assessment tool:*  [*https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/*](https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/) |

|  |  |
| --- | --- |
| 1. Implementing this Plan | Help, Guidance and *Example Answers* |
| 1. **Name of person responsible for implementing this plan.** | I will take responsibility for carrying out the actions required by this plan and report them to my supervisor as appropriate. |
| 1. **Frequency of review and/or updates of this plan.** | My supervisor and I will review it every 6 months and update if necessary |
| 1. **Actions required in order to implement this plan.** | Ask my supervisor to request research storage space for my project Set up a backup system. Test I can restore from my backup. Learn how to anonymise data for archival. |
| 1. **List any further information needed to carry out the actions above.** | * Where can you find this information? Who might you be able to ask?   e.g. DCC, DMPonline |

|  |
| --- |
| **Updated May 2021**  **This DMP template is based on:**   1. The UoB DMP template v5 2. Leiden University Data Management Plan template, v 4.2   Latest version of the template available on Zenodo: http://doi.org/10.5281/zenodo.3903267   1. Template: Evaluation Rubric for Data Management Plans (<https://scienceeurope.org/our-priorities/research-data/research-data-management/>) 2. Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2019). *Managing and sharing research data: a guide to good practice*. Sage. 3. <https://mantra.edina.ac.uk/datamanagementplans/> |