Data Management Plan (DMP)

This DMP template (1) is based on the FNR's template but contains additional information to guide you in writing your DMP. All templates and further information exists on the UL website and digitally on DMPonline. If you have any questions, please email ORBilu@uni.lu

General information	
Remember that the person reading your DMP may not know anything about your project and not be familiar with the discipline. It is important to be repetitive and descriptive. This is a narrative, contemplative exercise and not a tick box yes or no test. The reader expects you to be as reflective as possible on critical components from data curation to end preservation. At the same time, be concise and to the point ; they do not expect a novel. To prepare for documenting research projects and writing a DMP it is worth reading the knowledge articles on Atlas for Research Projects, File sharing solutions, and other relevant articles on the SIU knowledge portal.	
Administrative information	Provide information such as the name of the applicant, project number (if applicable), funding programme and version of DMP.
1 Data description and collection or re-use of existing data	
What the funders will be looking for is that you have described your data well, the type of data you will be using, collecting, creating, re-using. It is worth dividing your datasets into rows if you have lots of different data in your project. Then you describe each dataset separately. Where did the data come from? How will you collect them and what do you need to use and analyse the data. They also want you to explain why you do not re-use data, what were the reasons?	
1a How will new data be collected or produced and/or how will existing data be re-used?	A Explain which methodologies or software will be used if new data are collected or produced. B State any constraints on re-use of existing data if there are any. Explain how data provenance will be documented. C Briefly state the reasons if the re-use of any existing data sources has been considered but discarded.
Guidance	▪ Are the datasets **new** from this project? Were the data **re-used** from somewhere else? Describe where you **sourced** the data and if you sought **permission**. Give URLs, reference collections/databases...etc. ▪ If you are not re-using data, the explanation may be that this is not common practice in the field, or the data are not suitable. Provide context: why are you not reusing data. ▪ Describe the constraints. Constraints may mean data with a license. What are the conditions for re-using data? Check the license, you could ask your research facilitator or PAKTTo for help concerning licenses. Data license terms should be described, e.g. open source or other (https://opensource.org/licenses). ▪ Data collection/use methodologies/software, describe these as detailed as possible. Specifically;: • Experimentalists: what machine/equipment/protocol (own or commercial).

	If you use commercial software, is it interoperable? If it is your own software, describe the license terms. Is it open-source software? Detail the license. Worth reading: https://opensource.org/licenses • Surveys: explain how data will be collected (methodologies). • If you are a mathematician, 'data' may mean your mathematical formulas, explain how you produce these 'data'. • Data might also be related to artefacts or documents (if you are a historian, for example).
1b What data will be collected or produced?	A. Give details on the kind of data: for example, numeric (databases, spreadsheets), textual (documents), image, audio, video, or mixed media. B. Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example csv, xlsx). C. Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used. D. Give preference to open and standard formats as they facilitate sharing and long-term re-use of data. E. Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).
Guidance	Specifically; • Give a list with the type of 'data' in your project (formulas, qualitative data, photographs, instrument readings...)? • Open and standard formats means in practice: .txt (not .doc) or .csv (not .xls) files. Which formats are standard and robust over time?

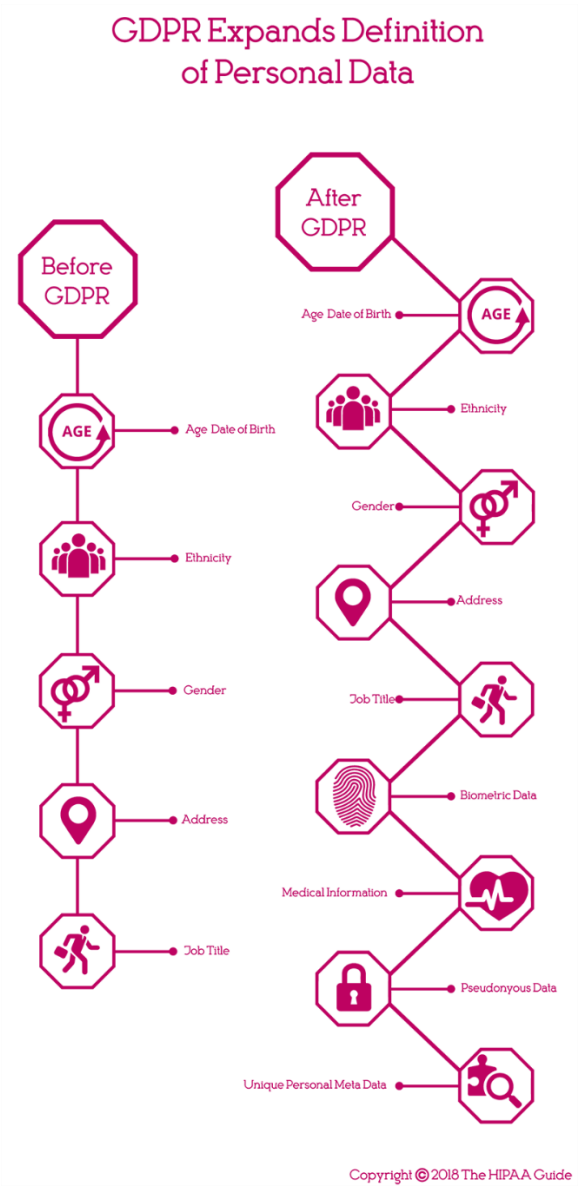
2 Documentation and data quality	
Funders will expect you to document your research data, explain how provenance will be documented, describe best practices; where you will store data, how will you name the files (name of project, producer, date), will the filename describe or identify sample or test run etc.? Describe the *logbook* of the project, e.g., electronic, hard copy, which details all conditions, samples, dates, experiments, test runs, etc. Describe that (experimental) conditions are always stored in the header of the file (if this is the case). Perhaps there are README files associated with the data also?	
2a What metadata and documentation will accompany the data?	A. Indicate which metadata will be provided to help others identify and discover the data. B. Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used. C. Use community metadata standards where these are in place. D. Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well- ordered research data will be easier to find, understand, and re-use. E. Consider what other documentation is needed to enable re-use. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on. F. Consider how this information will be captured and where it will be recorded, for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.
Guidance	▪ Metadata is 'data about data', e.g. information that describes and explains data, and their structure, in various document types. In databases (e.g ORBilu), there are standard field headers that organise information that you input so it becomes interoperable with other tools. Read more about metadata (external site). ▪ What approach will you use to capture information to keep the data understandable and usable? Perhaps you will create a notebook, README.txt files, codebooks etc. where information is recorded? ▪ Describe who will use the data and why. Data may be published in a graph or a figure, you can describe that. What do the data look like when it is shared? ▪ What type of data will be relevant to the community? Use the logic of your field. You can describe what data will remain on UL servers and how it may be accessible through metadata (to discover the data). ▪ (Hint, you can create an ORBilu metadata record for your data, so that it is discoverable by the public, even if the data themselves are not published. If you publish the data on a repository, e.g. Zenodo, you can add its URL to the ORBilu record.) ▪ Not all data should be shared, it may not be valuable. Describe what data you will share and not share, describe the non-value, e.g. non-userfriendly raw data etc. In material sciences most experimental data are not shared. ▪ For shared data, indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI, Dublin Core) will be used. Sometimes, software contains these already (worth checking). Use community metadata standards where these are in place. ▪ Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent,

	well-ordered research data will be easier to find, understand, and re-used.
2b What data quality control measures will be used?	A. Explain how the consistency and quality of data collection will be controlled and documented. B. 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.
Guidance	• It is important to address data quality, e.g. how will you safeguard against inaccuracies, inconsistencies (formats), systematic and human errors? How can the data be trusted as complete and consistent? How are your research methods to collect data valid? There may be certain standards in your field, for specific software, or methodologies and tools used. Perhaps there is a data governance framework? • 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. • Be precise but concise about each process. If there are no standard calibration, etc. you can describe this to say that you compare to benchmarking experiments and typical validation tests done in your field.

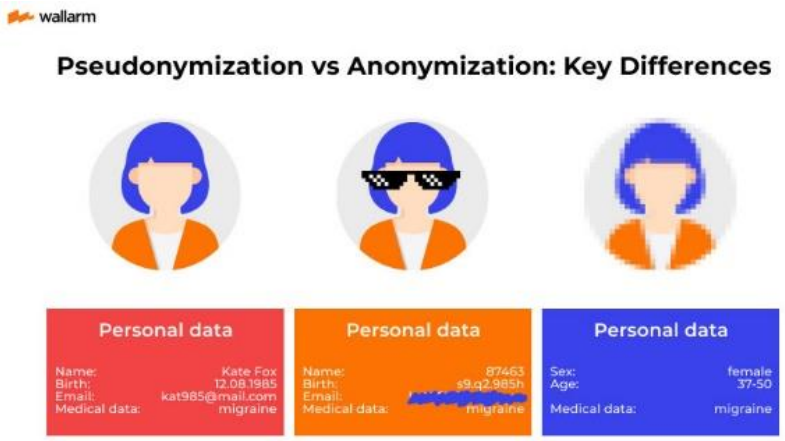
3 Storage and backup during the research process	
In this section you should describe your plan on how to store and backup both data and metadata. You should not store data on local PCs, and if you must, describe how you will back these data up regularly onto the servers or another PC. Think about if you have sufficient storage and backup capacity during the project. In particular show that you have an understanding of security and what data might be sensitive and how you deal with those.	
3a How will data and metadata be stored and backed up during the research?	A. Describe where the data will be stored and 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. B. 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.
Guidance	▪ You can describe Atlas as the secure server for the University that is regularly maintained and backed up. ▪ You can describe if you use SharePoint with the team to share data via the server. ▪ If you use external hard disks, etc., describe why and how you protect them. Hard disks are easy to lose or damage; rethink your strategy; they are not recommended! ▪ What backup takes place and how frequently?
3b How will data securityand protection of sensitive data be taken care of during the research?	A. Explain how the data will be recovered in the event of an incident. B. Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships. C. Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed. D. Explain which institutional data protection policies are in place.
Guidance	• You can describe that there is minimal risk associated with files stored on Atlas or other University servers. • Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships (onedrive is an option at the moment for UL). • Describe the research group access and if select data are accessed by specific researchers only. • Consider the difference between sharing and access of raw data and that of the results of data. Write how data are shared during the project, then what you expect after the project has finished, will you share via Zenodo? (each community might have a preferred one). Hint, you can make a metadata record on ORBilu for data that can be requested or made public, that way your data and publications are all in one repository. • Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed. Discuss with the DPO (data protection), legal department (if there are contracts in your project) and with PAKTTo on intellectual property, 'secret' data etc. You

can also speak to your data steward/manager (if your faculty has one) or your research facilitator, if you are uncertain of anything related to the project data.

- Remind yourself of GDPR regulations and what definitions are in place for anonymised data and pseudonymised data.
- Think about password protection on files that contain sensitive data, and where these passwords are recorded.
- Explain which institutional data protection policies are in place. Refer to UL guidelines for help with writing about data protection policies at the University.

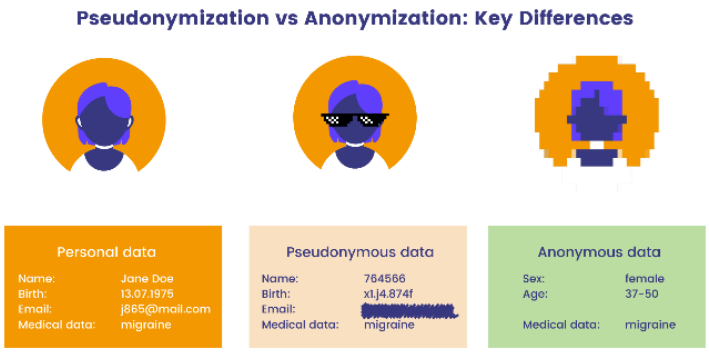


- Source https://www.hipaaguide.net/gdpr-for-dummies/



- Source https://www.wallarm.com/what/pseudonymization

- Source https://dataprivacymanager.net/pseudonymization-according-to-the-gdpr/

4 Legal and ethical requirements, codes of conduct	
This is a huge and complex topic, and if you are uncertain, seek advise from the DPO, PAKTTo, the Ethics Committee, and your Research Facilitator. You need to describe how you will process personal data, if at all. When you define the data you will collect/create, if any of these are personal, you are indeed 'processing' them even by simply collecting. How will you deal with intellectual property, are any of these data to be protected by a patent? Speak to PAKTTo about ownership of data, protection of data, and partnerships. Any ethical concerns speak to the the Ethics Committee. By filling in the ethics application you will be able to identify data that have ethical implications.	
4a If personal data are processed, How will compliance with legislation on personal data and security be ensured?	A. 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). • Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible). • Consider encryption which is seen as a special case of pseudonymisation. • Explain whether there is a managed access procedure in place for authorised users of personal data.
Guidance	Source: Image of anonymisation and pseudonymisation  • Source https://www.wallarm.com/what/pseudonymization ▪ Ensure that when dealing with personal data, data protection laws (for example GDPR) are complied with. ▪ Informed consent means gaining permission from research subject to collect and/or share their data – there are standard forms at UL that you can use. ▪ Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data). Anonymisation means that there is no trace of the person from the available data. ▪ Consider all potential access to any data that may reveal the identity of a person and remove the possibilities to such eventualities. ▪ Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible). Pseudonymisation

	means the use of a key (e.g. a code or similar) that identifies a person, for example an Id number. Describe the protection of the individual with the use of the pseudonymisation key. ▪ Consider encryption which is seen as a special case of pseudonymisation. ▪ Explain whether there is a managed access procedure in place for authorised users of personal data. ▪ You can describe how you minimise the access to sensitive personal data within the project team, by perhaps dividing tasks so that only one person can access one part and never the full data. ▪ How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable? ▪ Explain who will be the owner of the data, meaning who will have the rights to control access. Normally the PI is the owner of the data, and the research group has access, who else may have access, administrators, IT personnel, research facilitators…? ▪ 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.
4b How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?	• Explain who will be the owner of the data, meaning who will have the rights to control access: ○ 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. ○ 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. • Indicate whether intellectual property rights are affected. If so, explain which and how will they be dealt with. • Indicate whether there are any restrictions on the re-use of third-party data.
Guidance	▪ Examine the project contracts with collaborators to see what has been agreed with regards to data. ▪ Create a data project management team in the collaboration to discuss data management frequently. ▪ Draw up a data management framework. ▪ Use your DMP to guide you in data management.
4c What ethical issues and codes of conduct are there, and how will they be considered?	A. Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning. B. Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics a. committee) is required for data collection in the research project.
Guidance	▪ You can refer to codes such as those of membership bodies, e.g. Association of Social Anthropologists. ▪ Consider all aspects of ethics in research.

5 Data sharing and long-term preservation	
This is thinking beyond project, unless you shared data early (before project end). You can start thinking about this early on in the project and update the DMP annually as you know more about your data. You can refer to the UL archivist (coming soon) at UL for questions on preservation. You can contact the National Archives for questions on preservations too.	
5a How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?	A. 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). B. Outline the plan for data preservation and give information on how long the data will be retained at UL or if transferred somewhere, e.g. National Archives or another organisation. C. 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. D. Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. E. Explain what action will be taken to overcome or to minimise restrictions.
Guidance	▪ Explain data accessibility within the context of your field. If no data will be shared explain why. ▪ If some data can be shared define who and how and if there will be a license. ▪ Describe the data repository that you may use to share data. (Zenodo is a free repository hosted by CERN, and is useful to consider.) ▪ Outline the plan for data preservation and give information on how long the data will be retained. Consider long term storage costs early on in the project (this should have been agreed at funding application stage also). ▪ Describe the repository conditions, you can get these from the repository website. For the University we keep data on Atlas for 10 years. You can check with SIU if you require special storage conditions. ▪ Remember to describe if data may be used in other future projects. (Remember that this should have been outline in the ethics application also.) ▪ 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.
5b How will data for preservation be selected, and where data will be preserved long-term?	A. Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes. B. Indicate how it will be decided what data to keep. Describe the data to be a. preserved long-term. C. Explain the foreseeable research uses (and/ or users) for the data. D. 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 a. metadata standards, and costs involved) have been checked.
Guidance	▪ If you have any agreements that describe any of these preservation plans, refer to it. ▪ Why is specific data selected for long-term preservation, are they of special value, why and how do you decide that within the team? ▪ Refer to the repository you chose once more. ▪ Explain the foreseeable research uses (and/ or users) for the data. ▪ You can describe the specific value of that data to future users. ▪ 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.
5c What methods or software tools are needed to access and use data?	A. Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data. B. Indicate whether data will be shared via a repository requests handled directly, or whether another mechanism will be used?
5d How will the application of a unique and persistent identifier (DOI) to each data set be ensured?	A. Explain how the data might be re-used in other contexts. Persistent identifiers should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and re-use. B. Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.
Guidance	• Check with colleagues what repositories they use in your field. • NOTE: ORBilu is not a repository for data, but you can create a metadata record on ORBilu for a dataset, which will give you a persistent link (handle) that you can use in documents/publications. Your ORBilu record can describe the dataset, its accessibility, location and how to request it etc.

6 Data management, responsibilities and resources	
6a Who will be responsible for data management?	A. Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible (if agreed). B. For collaborative projects, explain the co-ordination of data management responsibilities across partners. C. Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised. D. Consider regular updates of the DMP.
6b What resources will be dedicated to data management and ensuring that data will be FAIR?	A. Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. B. 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. C. Indicate whether additional resources will be needed to prepare data for D. deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.

The questions are based on the FNR DMP template.
Guidance notes to DMP template
University of Luxembourg, 2024