NARODOWE CENTRUM NAUKI

Guidelines for applicants to complete the Data Management Plan form in the NCN proposal for calls:

- OPUS 19
- PRELUDIUM 19
- POLS
- EXPRESS CALL TO FUND RESEARCH ON COVID-19

NATIONAL SCIENCE CENTRE
POLAND

Guidelines for applicants to complete the Data Management Plan form in the proposal

In this section, the NCN understands 'data' to be both collected, unprocessed data as well as analyzed, generated data, other than scientific publications. Under this all forms are conceivable; digital and non-digital (for example samples, completed questionnaires, sound recordings, etc.).

Consider your DMP as a part of your research plan. DMP complements your research plan with a description of the technical management of your data. The NCN recognises that some projects will not generate, re-use or analyse research data and similar materials. In these cases, a short explanation is required.

For the completion of the section please contact the library/intended repository/ICT Department of your institute or university. They can help you with the completion of the data section.

	Questions	**Help text**
1.	**Data description and collection or re-use of existing data**	
	1.1 How will new data be collected or produced and/or how will existing data be re-used? Questions you might want to consider: – What standards, methodologies or software will be used if new data are collected or produced? – Which existing data (yours or third party) will you re-use? – How data provenance will be documented? – How will you organize your files and handle versioning? – Do analogue or paper-based research data (maps, photographs, text) need to be digitised to increase their potential for sharing?	Explain how the data will be collected, or produced. Also, mentioned any existing data that will be (re-)used. Briefly describe what types of data you are collecting or producing. Also explain what kinds of already existing data you will use. For example, the types of texts, images, photographs, measurements, statistics, physical samples or codes. Describe how you plan to control and document the consistency and quality of the collected data: calibration processes, repeated samples or measurements, data recording and capture standards, usage of controlled vocabularies, data entry validation, data peer review, etc.
	1.2 What data (for example the kinds, formats, and volumes) will be collected or produced?	The descriptions should include the type, format and content of each dataset.

	Questions you might want to consider: – What type, format and volume of data will you collect, generate or reuse?	Furthermore, provide an estimation of the volume of the generated data sets. Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (i.e. pdf. xls.) Give preference to open and standard formats.
2.	**Documentation and data quality**	
	2.1 What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data? Questions you might want to consider: – What information is required for users (computer or human) to read and interpret the data in the future? – Is the data machine-readable? – How will you generate this documentation? – What community standards (if any) will be used to annotate the (meta)data? – What international standards or schemes (i.e. Dublin Core, DDI) will be used to structure metadata?[1]	Indicate which metadata will be provided to help others identify and discover the data. This may include information on: the title of the files, sources of the data, author ID number (e.g. ORCID), formats methodology used to collect the data, definitions of variables, units of measurement. Indicate how the data will be organised during project, mentioning for example conventions, version, and folder structure. Consider how this information will be captured and where it will be recorded i.e. in a database with links to each item, README files, code books etc.
	2.2 What data quality control measures will be used? Questions you might want to consider: – How the data collection, analysis and processing methods used may affect the quality of data? – How measurement error and bias will be eliminated?	Illustrate that data possess high quality attributes. Are data collection and analysis methods documented? Indicate the existing mechanisms to prevent unauthorized changes in the institution. Describe how/when

[1] The Digital Curation Centre (DCC) maintains a list of widely used disciplinary metadata standards. Researchers are advised to consult this list. http://www.dcc.ac.uk/resources/metadata-standards.

	– How you will minimise the risks related to data accuracy?	internal data quality assessments will be implemented. Explain whether quantitative data needs to be cleaned.
3.	**Storage and backup during the research process**	
	3.1 How will data and metadata be stored and backed up during the research process? Questions you might want to consider: – What is your storage capacity and where will the data be stored? – What are the back-up procedures? – Are special measures needed to transfer data from mobile devices, from fieldwork sites or from home equipment to a central work server?	Please mention what the needs are in terms of data storage and where the data will be stored. For long-term storage decide which data will be kept, which storage volume this represents and how long data will be stored and preserved. Please consider that data storage on laptops or hard drives, for example, is risky. Storage through IT teams is safer. Please specify your back-up procedure (frequency of updates, responsibilities, automatic/manual process, security measures, etc.). Consider who will be responsible for backup and recovery. If there are several researchers involved, create a plan with your collaborators and ensure safe transfer between participants.
	3.2 How will data security and protection of sensitive data be taken care of during the research? Questions you might want to consider: – How the data will be recovered in the event of an incident? – Who will have an access to the data during the research and how access to data will be controlled, especially in collaborative partnerships?	If external services are asked for storage, it is important that this does not conflict with the policy of each entity involved in the project, especially concerning the issue of sensitive data. Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, dual-use data. Explain which institutional data protection policies are in place.
4.	**Legal requirements, codes of conduct**	

	4.1 If personal data are processed, how will compliance with legislation on personal data and on data security be ensured? Questions you might want to consider: – Do you need to use anonymisation throughout a data collection? – Do you need to remove identifying information or conceal the identity of participants (e.g. using pseudonymisation) before data can be shared?	Ensure that when dealing with personal data protection laws (i.e. GDPR) are complied with gain informed consent for preservation and sharing personal data. Consider anonymisation or pseudonymisation of personal data. Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data). Explain whether there is a managed access procedure in place for authorised users of personal data.
	4.2 How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable? Questions you might want to consider: – Who will be the owner of the data? – Which licenses will be applied to the data? – What restrictions apply to the reuse of third-party data? – Do you need to seek copyright clearance before sharing data?	Outline the owners of the copyright and Intellectual Property Right (IPR) of all data that will be collected and generated, including the licence(s). For consortia, an IPR ownership agreement might be necessary. Furthermore, clarify whether there are any restrictions on the re-use of third-party data. Indicate whether intellectual property rights (for example Database Directive 96/9/EC at March 11th 1996, *sui generis* rights) are affected.
5.	**Data sharing and long-term preservation**	
	5.1 How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons? Questions you might want to consider: – How will potential users find out about your data? – For how long will the data be stored? – Are there any barriers and constraints to making the research data fully or	Data have to be shared as soon as possible, but at the latest at the time of publication of the respective scientific output. Please, also consider how the reuse of your data will be valued and acknowledged by other researchers. Explain when the data will be

	partially accessible? – Will journal publishers require deposit of data supporting article findings? – Do you need to ask participants for their consent for data to be shared?	made available. Justify the retention period for data storage[2]. Indicate the expected timely release. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents. Consider whether a non-disclosure agreement would give sufficient protection for confidential data.
	5.2 How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)? Questions you might want to consider: – What data must be retained or destroyed for contractual, legal, or regulatory purposes? – How it will be decided what data to keep? – What procedure would be used to select data to be preserved? – What repository will you be using? Is this repository conform to the FAIR Data Principles[3]? – Does the institution provide regular data backup or not?	Consider how and on which repository[4] the data will be made available. Consider the cost of data deposit and storage space for long-term storage. Estimate how much data storage space is needed for the entire duration of the project and how much you need after it ends. Please, explain whether you choose digital repository maintained by a non-profit organisation?

[2] Raw and processed data must be stored for a period appropriate for the discipline and methodology at issue. NCN considers a minimal period of 10 years reasonable.

[3] The FAIR Data Principles define a range of qualities a published dataset should have in order to be Findable, Accessible, Interoperable and Reusable (see Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, doi:10.1038/sdata.2016.18) and https://www.force11.org/fairprinciples.

[4] There are a number of international certification schemes, which determine the trustworthiness of data repositories. Of these the international Data Seal of Approval certified by Core trust Seal is the most basic set of criteria. Trusted Digital Repositories with a quality mark include repositories with a Data Seal of Approval, DIN-31644-, ISO-16363- or WDS/ICSU certification. An overview of existing repositories with Data Seal of Approval can be found in this list of repositories. Other useful listings of repositories include: Registry of Research Data Repositories https://www.re3data.org/, some of them like Zenodo, an OpenAIRE and CERN allow researchers to deposit both publications and data, while providing tools to link them. It is always recommended to refer to broadly recognised discipline-specific or certified repositories in the first place. In cases where no such a repository can be identified for selection of trustworthy repository please use criteria listed in *Practical Guide to the international alignment of research data management*, https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf.

	5.3 What methods or software tools will be needed to access and use the data? Questions you might want to consider: – Do data need to be converted to a standard or open format with long-term validity for long-term preservation? – Is additional equipment or software needed for scanning or conversion? – What mechanism will be used for data sharing; e.g. request handled directly, repository?	The methods applied to data sharing will depend on several factors such as type, size, complexity and sensitivity of data. Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data.
	5.4 How will the application of a unique and persistent identifier (for example: such as a Digital Object Identifier (DOI)) to each data set be ensured? Questions you might want to consider: – Will persistent identifier for the data be obtained? – Which existing persistent identifier will be used (e.g. Digital Object Identifiers, Accession Numbers)?	Explain how the data might be re-used in other contexts. Persistent identifier 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.
6.	**Data management responsibilities and resources**	
	6.1 Who will be responsible for data management (i.e. data steward)? Questions you might want to consider: – Does your institution have staff responsible for data management during the project, i.e. for initial collection, archiving and data curation? If NO, who is responsible for data archiving and long-term preservation? If YES what is his/her position in the institution? – What is the role of PI in data management process?	It has to be distinguished between Data owner and Data steward. Data steward is a data quality (DQ) expert who is responsible for data assessment (corrective measures) but he/she is not in charge of correcting data themselves. Outline the roles and responsibilities for data management/stewardship activities. Indicate who is responsible for implementing the DMP, and ensuring it is reviewed and revised. For collaboration project, explain the co-ordination of data management responsibilities across partners.

	6.2 What resources will be dedicated to data management and ensuring that data will be FAIR[5]? Questions you might want to consider: – Do you need extra resources to manage data, such as people, time, hardware or software? – What are the costs for making data FAIR in your project? – How will these costs be covered?	Explain how the necessary resources to prepare data for sharing/curation have been costed in. Indicate if the additional resources will be needed to prepare data for deposit. If yes, please explain how much is needed and how such costs will be covered. Consider the cost of data deposit and/or longer-term storage – find out from data centre/repository whether charges apply.

Additional information:

If you submit an application in an international competition organized by the NCN based on the procedure of a leading agency, in which the leading agency is a foreign partner, regardless of whether this agency requires submission of a data management plan or not, you are still obliged to supplement it in the NCN application form submitted in the ZSUN/OSF system. The content of the plan applies only to research data that will be created or will be reused during the implementation of the project by the Polish research team.

If you submit an application in an international competition organized by the NCN in multilateral cooperation, regardless of whether you need to submit a data management plan at international level, you still have to complete it in the NCN application form submitted in the ZSUN/OSF system. The content of the plan applies only to research data that will be created or will be reused during the implementation of the project by the Polish research team.

DMPs are very individual. They can be of various types and their composition can differ. The examples provided by the Digital Curation Centre (UK) show this diversity.

[5] There are a number of self-assessment tools that might be helpful to assess the FAIRness of your data; i.e. https://www.ands-nectar-rds.org.au/fair-tool.