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. 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.

legend:

italics applied to the text, marked yellow – information to be read and completed if necessary
red text font - text to choose from depending on the specifics of the project
black text font - information to be copied and possibly corrected depending on the project

	Questions	**Help text**	**Sample text prepared by IRziBŻ PAN in Olsztyn**
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? • What quality assurance processes will you use? • 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?	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.	During the implementation of the project, two types of data will be generated to further processing: a) automatic measurement data resulting from the scientific research equipment, which result is an electronic data set, *(- please enter the device name and what types of files they generate* *- if a file naming algorithm is used please state how it works)* b) 'non-automatic' measurement data resulting from tests / experiments carried out personally / manually by scientist - data documented manually and then entered into a spreadsheet (MS Excel * .xlsx format; or other). The acquired data will subsequently be subject to statistical processing, ..., analysis. During the project implementation, data that will be generated include numerical data, text documents, photographs, content of databases, interaction analysis, laboratory reports, methodological description. *Project do not plan usage of data generated previously is no planned, but during generating new data they are planned to be adjusted to for multiple usage according to FAIR rules. The whole data set will be collected and processed by the project leader.* OR *During project execution already generated data under license CC-BY or CC-BY-NC will be used (please check the license under which the data is available and enter*

			the appropriate license). During generating new data they are planned to be adjusted to for multiple usage according to FAIR rules. The whole data set will be collected and processed by the project leader.
	1.2 What data (for example the kinds, formats, and volumes) will be collected or produced? Questions you might want to consider: • What type, format and volume of data will you collect, generate or reuse?	The descriptions should include the type, format and content of each dataset. 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.	The following formats will be used during measurements and analysis: *TIFF, xrdml, opj, csv, txt, xlsx*. All data selected for long-term archiving and sharing, will be in suitable open format like: *txt, pdf, TIFF*. We expect to have *8 GB* of data in this 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.	During the project there is conducted documentation in paper (e.g. laboratory notebook) and electronic version. Each experiment has plan of experiment, procedure in paper and electronic version, or used protocol. In document description there is a title of experiment, data of preparation and execution, person responsible for, procedure description and publication connected with and citation of the methodology. We will use README files to document data background and each stage of data generation stage / processing. The description of catalogues scheme is as follows: Project name/ Experiment name/Date/File_name *(please enter how you will describe the folder)* *If you can deposit your data in what repositories, databases, data banks (e.g. nucleotide sequences, microarrays) etc. - please provide them.* Chosen data will be shared via open repository (provide its name) togheter with metadata standards: (eg. DataCite, DablinCore etc.) please check in your browser: https://www.re3data.org/, what standards the selected repository has). If your repository allows it.
	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 internal data quality assessments will be implemented.	The methods and data acquisition are given in the documentation regarding the given study / experiment. The quality of the data and its ongoing control will take place at several levels: - detailed planning of each experiment, - keeping documentation of the experiment, enabling to trace the history at any time,

[1] 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?	Explain whether quantitative data needs to be cleaned.	- use of previously created experimental protocols, thus ensuring constant, controlled experimental conditions, - performing preliminary statistical analyzes of obtained data to detect anomalies, possible errors/mistakes, necessity of repetition the experiment, - use of appropriate control groups. Appropriate effort will be taken when characterizing data according to FAIR rues. Raw data will be in majority generated and collected automatically with measuring equipment that will be properly calibrated. *At this point, you should look at the project individually and provide project-specific activities.*
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? • Do analogue or paper-based research data (maps, photographs, text) need to be digitised to increase their potential for sharing?	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.	The data will be stored on local computers, external hard drives, and on the local server of IAR&FR of PAS (long-term storage). Data backup on local computer is created automatically using research data management system of IAR&FR PAS that currently is under development. Long-term storage of data takes place on IAR&FR PAS servers, on a separate disk space, in consultation with the IT Systems Administrator IRZiBŻ PAN (backup of this data is performed automatically). Access to data on the server is possible only from computers registered in the internal network, it is set individually using a login and password. *If scientists use other methods of data storage (cloud) and their archiving - it should be written about it.*
	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.	IAR&FR of PAS acts in accordance with Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46 / EC (hereinafter referred to as " RODO "). In IAR&FR of PAS has been evolved The Book of Personal Data Protection Policies referred to in art. 24 paragraph 2 GDPR, which was implemented by the Ordinace No. 3/10/2021 of the Director of IAR&FR of PAS in Olsztyn from 28.10.2021. To secure backups the research data management system of IAR&FR PAS that currently is under development will be used. Only authorized project staff will have access to data. Data sharing will be conducted via IAR&FR PAS infractructure and

			protected by password or *in another way (describe the method of sharing).* Backups of all the data will be done through the whole project. *when sensitive data are processed:* Sensitive data that allows identification will be anonymized/ or pseudonymized*(please specify what will be done with the sensitive data). There i salso planned to create regulations governing issues of data access – list of authorised staff, list of staff with the access to the room where data are stored. Data on paper will be locked using key (if they are).* *or when sensitive data is not processed:* The project will not process sensitive data.
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.	*Only applies to projects where personal data will be processed* Sensitive data will be stored in a locked room, to which authorized persons have access, such as PI. Each person, participating in examination, will be provided with information about the processing of personal data and acceptance of the rules. The data will be collected, processed only for the purposes of the project, encoded data of participants will be used in any publications. Identification codes given to study participants will be used in the analyzes.
	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, sui generis rights) are affected.	Intellectual property rights belong to IAR&FR of PAS. The rules for the division of intellectual property rights between the institute and the creator are described in the Ordinace No. 134/2012 of the Director of IAR&FR of PAS of 17.12.2012 regarding the introduction of regulations on the protection of intellectual property and knowledge transfer and the Regulations of the rules and procedures for commercialization of the results of scientific research, development works and know-how related to these results and the management of industrial property rights in IARFR PAS (Resolution No. 12 of the Scientific Council of the IARFR PAS of 26th February 2015). *In the case of projects implemented in concert:* If a project is implemented by several partners, the owner of the intellectual property rights to the products / results created as part of the Project will be the Party who created the product / result. In the case of products / results created as part of a project jointly created by the Parties, intellectual property rights will be vested in parts established on the basis of separate agreements, taking into account the intellectual, financial and material contribution of the Parties to the

			creation of these products / results. These rules are included in Agreement signed by Partners.
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 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?	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 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.	Potential users will learn about the research carried out after publication in scientific journal. Additional research data will be made available through the journal website. The sensitive data not be shared, the shared data will not require the permission of the study participants. Access to (insensitive) data will also be possible upon an individual request to the PI and the person representing IAR&FR of PAS. Data will be stored in repository (provide its name) as long as it is technically possible.
	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?	Consider how and on which repository[4] the data will be made available. Outline the plan for data preservation and give information on how long the data will be retained. 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. Please, explain whether you choose digital repository maintained by a non-profit organisation?	Data will be chosen according to their scientific merit. Data that will be useful for other investigators will be shared using repository (provide its name), that respects FAIR rules. Data with raw material will be stored on local computer among author's data sets as well as on IAR&FR PAS research data management infrastructure what includes providing automatic backups.

[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.

[4] There are a number of international certification schemes, which determine the trustworthiness of data repositories. Of these the international Data Seal of Approval 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

	• 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?		
	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.	The data will be made available in response to an individual request. The data will be stored in the source format and after processing, e.g. statistical and. In some cases, specialized software will be required to read data (*specify which for the given project / study*). IRZiBŻ is equipped with hardware and software enabling scanning and data conversion to the appropriate format. For data sharing the infrastructure of repository (provide its name) will be used. Shared data will be uploaded in the open formats what will guarantee that viewers would not need to use specialistic software. Data will be shared in the repository and stored according to the agreement.
	5.4 How will the application of a unique and persistent identifier (such us 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.	Scientific articles will be assigned a DOI number identifying a uniquely the scientific article. Shared data sets will be assigned a DOI number or other identification number accordingly to the repository.
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: • What is the role of data steward in your institution? • What is his/her position in the institution?	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	Person that will be responsible for data management is project leader. In case that project leader is no longer an employee of the IAR&FR PAS, Director designated a Data Steward who will take the responsibility for at least 10 years.

[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).

		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: • What are the costs for making data FAIR in your project? • How will these be covered?	Explain how the necessary resources (for example time) 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.	The project does not provide for separate financial resources for data management and ensuring the possibility of finding, accessing, interoperability and re-using data and archiving. IRZiBŻ PAN provides the infrastructure necessary for data management and storage, and Data Steward has been established.

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.