Instructions for Completing the Lendület 2026 Research Data Management Plan (DMP)

These instructions are designed to guide you through the completion of the Research Data Management Plan form required for the Lendület 2026 grant application. A well-crafted DMP demonstrates responsible and effective data management practices, crucial for the successful research project. Please read these instructions carefully in conjunction with the provided form, keeping in mind the principles of FAIR data (Findable, Accessible, Interoperable, Reusable) and the guiding principle of making data "as open as possible, as closed as necessary." Remember that a DMP is a "living" document, that is updated as the project evolves. The creation of a DMP also helps to plan the necessary resources, verify the research methodology, and potentially make data available for future research

General Guidance:

As a general and useful piece of advice, it is important to emphasize that the use of technical terminology related to a given scientific field is extremely important. The exploration of the large amount of data and files present in the digital space is now often carried out by search engines and other automated systems. Partially for that reason, it can be stated that proper metadata annotation is perhaps the most important element or guarantee of compliance with each of the FAIR principles. If a data package uses appropriate vocabularies and ontologies, it is more than likely that the work will have a greater chance of meeting the individual FAIR criteria. Similarly, it is important to highlight the use of appropriate repositories (recognized and frequently used within the given scientific field), which, complemented by the aforementioned ontologies and keywords, ensure the implementation of the FAIR principles.

There may often be overlaps among the different aspects, so it is important to note that if a topic or piece of information is relevant to several different sections of the DMP, we should not feel that repeating the information makes it redundant. Beyond the present proposal, one of the most important roles of the DMP is to provide support to the research team members so that everyone clearly understands what is happening to the data. A DMP is nothing more than a plan describing how the data involved in the research will be managed in a FAIR manner and ensuring that throughout the process all team members understand and apply it consistently in practice.

The DMP should be considered a living document, subject to updates as the project evolves. Consistent version control (including dates) is essential. Tailor your plan to the specific standards and expectations of your research field. Clarity and conciseness are paramount; provide sufficient detail without unnecessary jargon. Be realistic about resource availability and address all ethical considerations related to data collection, processing, sharing, and storage. The overarching philosophy should be to make your data openly available whenever feasible, while justifying any necessary restrictions. The DMP aims to help the funding body and host institution plan necessary resources, assess research methodology, verify research, and identify data reusable for future research.

Important Note: If your research does *not* generate data (e.g., purely theoretical mathematics), please state this clearly in the SUMMARY section, and you may omit the subsequent sections. Please note that handwritten physical notes also can be useful and considered as research data if you can digitalize it. Remember that the specific interpretation

of each section should be tailored to your discipline. Different fields may have specialized repositories and standards that are best suited for your data. The examples provided in this document are just short generalistic illustrations for the kind of data that is expected for a given section, please be thorough when it comes to your specific project.

SUMMARY

Begin by providing a concise overview of your data management approach. This section should encapsulate the purpose of data collection/generation, its relation to the project's objectives, a unique dataset reference and name, the types and formats of data involved, the origin and expected size of the data, details regarding the reuse of existing data (if applicable), and an explanation of the data's potential utility and target audience. *Brief example:* "This project aims to collect survey data on public attitudes towards climate change. The data will be used to assess public awareness and inform policy recommendations. The dataset will be titled 'ClimateChangeAttitudes_Survey2026' and will consist of approximately 5,000 responses in CSV format. We will also utilize existing census data for demographic analysis." Remember to consider *who* would benefit from access to your data and *how* they might use it.

1. MAKING DATA FINDABLE

This section focuses on ensuring discoverability. Provide a detailed description of your dataset(s) and the metadata you will create. Metadata is "data about data" and is crucial for understanding and reusing your datasets. Utilize persistent and unique identifiers (PIDs), such as Digital Object Identifiers (DOIs), to facilitate identification and access. Detail how metadata will be provisioned to ensure discoverability, alongside established naming conventions and a considered approach to search keywords and versioning. *Brief example:* "Metadata will conform to the Dublin Core standard and will include fields such as title, author, keywords, abstract, date of creation, and geographical location. We will use descriptive file names following the format 'YYYYMMDD_Subject_Version.csv'. The DOI will be requested through our institutional repository upon completion of the project." Standardizing your naming conventions and using relevant keywords will significantly improve discoverability. Further important (non-specialized) PIDs to consider:

ORCID (free, unique, persistent identifier for individuals to use as they engage in research, scholarship, and innovation activities)

ROR (Research Organization Registry - a community-led dataset that aims to provide a persistent identifier for every research organization in the world)

Furthermore it is crucial that when you decide to upload your work onto an archive or repository, to always mark these PIDs there.

2. MAKING DATA OPENLY ACCESSIBLE

Clearly indicate which datasets will be made openly available and provide justification for any datasets that will remain closed due to legal, contractual, or voluntary restrictions. It is also important to emphasize that regardless of the level of restrictions, the precise and detailed

specification of metadata is indispensable, for example, in the case of depositing data in a repository. Explain how access will be granted, including any necessary software or methods, and specify the intended data deposition location (e.g., a research data repository like HUN-REN ARP, Zenodo, or a publications-specific repository). Detail how access will be managed in cases of restricted availability. *Brief example:* “All anonymized survey data and analysis scripts will be deposited in the HUN-REN ARP repository with a CC-BY license. Access will be open to all researchers. However, individual participant responses containing potentially identifying information will be kept confidential and will not be made publicly available due to ethical considerations." Remember, even if data cannot be fully open, consider providing access to metadata or aggregated results. Repositories like those mentioned offer long-term preservation and ensure your data remains accessible.

3. MAKING DATA INTEROPERABLE

Ensure your data can be combined and reused by identifying the standards and field-specific vocabularies you will employ to facilitate interdisciplinary data exchange and re-use among researchers, institutions, organizations, countries, etc. Adhere to open and standardized data formats and utilize open-source software applications whenever feasible to enhance data access and reuse. Consider how your data might integrate with other datasets from diverse origins. *Brief example:* "We will use the widely accepted ISO 8601 standard for dates and times. Quantitative data will be stored in CSV format. We will utilize controlled vocabularies from established ontologies in our field, such as the Gene Ontology, where appropriate. Our analysis scripts will be written in R, an open-source programming language." Utilizing widely accepted standards allows others to seamlessly integrate your data with their own, maximizing its impact. Consider the importance of rich metadata and utilizing metadata standards to facilitate machine readability and enhance data interpretability – to provide further insight it might be useful to write Readme files for the dataset.

4. INCREASE DATA RE-USE

Outline which data will remain reusable and for how long. Indicate if any embargo periods are planned (reasons, for how long). Explain how the data is licensed - choosing an appropriate license (like a Creative Commons license) clarifies the terms of use for others. Detail your data quality assurance procedures to ensure the data's reliability and validity. Confirm that your data can be utilized by third parties, particularly after project completion. *Brief example:* “All data will be made reusable for at least 5 years following the project’s completion. We will use a CC-BY license to allow for unrestricted reuse with attribution. Data quality will be ensured through rigorous data cleaning, validation, and documentation of all processing steps." The use of standardized licenses facilitates reuse and avoids legal ambiguity.

5. ALLOCATION OF RESOURCES AND DATA SECURITY

Provide an estimated budget for making your project data open access and potential costs of long-term data preservation. Consider costs associated with data storage during the research phase, repository fees, and the preparation of data for archiving. Explain how these costs will be covered. Describe your procedures for data backup and recovery, as well as the methods

for transferring and securely storing sensitive data in a repository for long-term preservation and curation. *Brief example:* "We have allocated €500 for data storage and repository fees. Data will be backed up daily to a secure off-site server. Sensitive data will be encrypted during transmission and storage." Think about data security (and its consts) from the outset, especially if you are dealing with sensitive or confidential information. There might be necessary expenses for softwares that you need to acquire.

Disclaimer: The researchers are ultimately responsible for addressing any ethical concerns related to data collection, processing, sharing, and storage throughout the project lifecycle.

Helpful Links:

- Lendület "Momentum"Programme: https://mta.hu/data/dokumentumok/lendulet/Lend%C3%BClet_2026/Lendulet_2026_call_for_application_.pdf
- HUN-REN Cloud Resources: https://researchdata.hu/en/guide-managing-research-data
- Re3data Repository Registry: https://www.re3data.org/
- DMP Tool: https://dmptool.org/

For any further questions, please contact the MTA KIK Open Access team at openaccess@konyvtar.mta.hu.