VolkswagenStiftung

Guidelines on Handling Research Data

Guidance for applicants, reviewers and project teams

Now more than ever, research data is both the foundation and the outcome of research. Research data comes in many forms, including raw data such as measurement data, laboratory values, audiovisual information, texts, survey data, etc. as well as secondary data such as software outputs, simulations, annotations, etc. – reflecting the diversity of disciplines, methods and research interests. Safeguarding research data is important for both the reproducibility of studies and as a basis for new research. In line with the FAIR principles (Findable – Accessible – Interoperable – Reusable), the Volkswagen Foundation supports the sustainable storage of research data and access pathways to this data that are as open as possible. Engaging with research data management in a structured way before a research project begins increases data quality, improves data security and makes the entire project more efficient in terms of time and resources.

These guidelines provide orientation for the description and assessment of research data management in projects for which funding is requested from the Volkswagen Foundation. They are based on Science Europe's research data management guidelines.

We recommend creating a data management plan (DMP) in the project planning phase. This helps to ensure that the work and resources required are planned realistically. The DMP should be updated as needed during the project and be considered a 'living document' (see Section 2). Section 1 covers only the part of the DMP that is described in the proposal and considered in the review of the proposal.

1 Description and Assessment of Research Data Management in the Proposal

The following sections correspond to the headings in the proposal's data management section and present guidance for applicants and reviewers side by side.

1.1 Data Description

Data management questions and guidance for applicants		Assessment guidance	
		Sufficiently addressed: The data management section…	**Insufficiently addressed: The data management section…**
What existing data is relevant to the project? Which of these will be used (source, terms of use)? If existing data cannot be used, explain why.	• Check whether data exists that could be used for the project. • If you reuse data, state its source, how it can be accessed and the licence or other terms of use. • If using existing data is not an option, briefly explain why.	• outlines which relevant data exists. • clearly states where the existing data comes from. • states, if existing data is used, how this data will be accessed and what restrictions apply to its use. • clearly explains, if applicable, why new data needs to be collected.	• does not state whether relevant data is available for reuse in the project. • provides little or no information on which data will be used and where it comes from. • if applicable, does not provide sufficient justification for collecting or generating new data.
What data (including data types and formats) will be collected or generated in the project? What data volume is to be expected?	• State which methods or software will be used to collect or generate new data. • Give details of the type of data, e.g. numerical (databases, spreadsheets), textual (documents), image, audio, video and name the file formats, e.g. pdf, csv, txt, xlsx, docx. • Give preference to open and standard formats as these facilitate the exchange and long-term reuse of data and justify the use of any other formats. • Provide information on the volume in terms of the storage space required (bytes) and/or the number of objects, files, rows and columns.	• clearly states how new data will be generated. The methods and software used are clearly named. • clearly describes or lists which data types will be generated and which file formats are associated with them. • provides a sound justification for choosing other file formats if these are <u>not</u> open and standard formats. • states the estimated data volume. • if applicable, clearly states that no new data will be generated in the project.	• provides little or no information on which data will be collected. • does not describe, or insufficiently describes, the data types generated in the project. • lists/describes data types without specifying their formats. • lists file formats without specifying the data types. • does not provide an estimate of the data volume.

1.2 Documentation, Legal and Ethical Aspects

Data management questions and guidance for applicants		Assessment guidance	
		Sufficiently addressed: The data management section…	**Insufficiently addressed: The data management section…**
What documentation and metadata will be created for the collected or generated data? How will discipline-specific standards be taken into account?	• Consider what documentation is needed for re-use. This may include information on data collection, analysis, definitions of variables, units of measurement, etc. • Consider how this information will be captured and where it will be stored (e.g. readme file, codebook). • State which metadata (and, where applicable, according to which standard or schema) will be created to make the data findable later. • Address discipline-specific practices and, where applicable, standards for organising and filing the data. For interdisciplinary projects, different standards may be named for the respective data types.	• clearly describes the documentation required for reusing the data and states where this information will be stored (e.g. readme file, codebook, lab notebook, database with links to the individual items). • clearly outlines the metadata, with reference to good practice (e.g. metadata standards).	• provides no information or only very vague references to documentation, without further explanation. • provides little or no information on the metadata that will accompany the data.
What ethical or legal aspects need to be considered when collecting and processing the data? Where applicable, how will the confidentiality of personal or otherwise sensitive data be ensured?	• Address discipline-specific practices and, where applicable, standards for collecting personal or other sensitive data (e.g. informed consent, protection of anonymity). • Describe which technical and organisational measures you will take to protect sensitive data (e.g. personal data or other sensitive information) during and after the project. • Consider third-party rights and deletion periods. • State whether your project requires approval from an ethics committee and, if so, how it will be obtained.	• clearly sets out how personal and other sensitive data will be handled responsibly. • clearly describes protective measures such as access and usage management and/or separately secured areas on the institution's own infrastructure. • states whether approval from an ethics committee is required and, if so, how it will be obtained.	• provides no information on whether ethically sensitive data will be collected and processed. • provides no or insufficient information on protective measures, even though it is foreseeable that sensitive data will be collected and/or processed. • provides no or insufficient information on whether approval from an ethics committee is required.

1.3 Archiving and Publication

Data management questions and guidance for applicants		Assessment guidance	
		Sufficiently addressed: The data management section…	**Insufficiently addressed: The data management section…**
Where and for how long will the data and metadata collected or generated in the project be archived after the end of the project?	• Outline the plan for archiving the data in line with good research practice (own infrastructure, archive, repository or similar) and state how long the data will be archived. • Where applicable, state any different retention periods for personal and sensitive data.	• provides information on how long the data will be archived (usually 10 years). • names the archiving location (own infrastructure, archive, repository), where applicable including the name of the service with which the data or metadata will be archived.	• provides little or no information on how, when, where and for how long the data will be archived.
At which point will the data be available to the research community, and under which licence? In which repository will the data be published? Where applicable, justify any restrictions or the decision not to publish.	• Explain when the data will be made available. • State the planned licence, e.g. CC BY. • Explain how the data will become findable and accessible (e.g. by depositing them in a trustworthy repository, indexing them in a data catalogue or using another mechanism). • Give preference to a discipline-specific repository. If no discipline-specific repository is used, explain why not. • When choosing a repository, make sure that persistent identifiers (PIDs) are assigned so that the data can be referenced and cited in the long term. • If you do not intend to publish the data or intend to make them accessible only with restrictions or only after an embargo period, explain the reasons (e.g. to enable publications, protect sensitive data or apply for patents) and the intended embargo period.	• states when the data will be made accessible (expected publication date) and under which licence. • names the repository and/or data catalogue in which the data is expected to be stored or listed. • explains whether and how persistent identifiers will be obtained. As a rule, recognised repositories assign PIDs. • if applicable, clearly explains why publication is restricted or not possible and who can access the data under which conditions. • explains, as far as possible, what measures will be taken to overcome or minimise restrictions on publication.	• provides little or no information on when data will be made accessible, or the explanations are inadequate or not technically feasible. • provides little or no information on the repository in which the data will be stored. • does not state whether (and which) PIDs the data will receive and gives no valid reason for not using them. • provides no clear information on whether individual datasets and/or collections will be assigned PIDs.

2 Recommendations for Handling Research Data During the Project

Good data management is essential for project success and for making data reusable. We recommend creating a complete data management plan (DMP) in the planning phase and keeping it updated as necessary throughout the project. The data management section of the proposal covers only part of this. The following overview lists the questions that a DMP should answer:

- **Data description:**
  Which existing data or collections will be reused?
  What data (types, formats, volume) will be collected or generated in the project, and how will it be processed?
- **Documentation and data quality:**
  What documentation and metadata will be created, and according to which standards?
  How will the data be organised (e.g. folder structure, version control), and how will its quality be assured?
  What software is needed to use the data?
- **Storage and backup during the project:**
  Where will the data be stored, and how will it be backed up and restored?
  Who will have access, and how will sensitive data be protected?
- **Legal and ethical aspects:**
  Which data protection and ethical requirements apply?
  How are copyright, personality rights, exploitation and ownership rights regulated, including for third-party data and between project partners?
- **Archiving and long-term accessibility:**
  Where and for how long will the data be archived?
  Which data will be made accessible, from when, under which licence and in which repository that assigns persistent identifiers (PIDs)?
- **Responsibilities and resources:**
  Who is responsible for the data and for updating the DMP during and after the project?
  What resources are required (e.g. staff time, storage, repository charges)?

The following also applies to funded projects:

The Foundation expects the use of the data generated or used in the project, as well as data ownership, access and exploitation rights during and after the project, to be governed between the project partners, e.g. in an agreement.