1. General
1.1. Topic
1.1.1. What is the main research question of the project?
1.1.2. Please give some keywords describing the research question.
1.1.3. Which research field(s) does this project belong to?
1.2. Project schedule
1.2.1. When does the project start?
1.2.2. When does the project end?
1.3. Project partners
1.3.1. Which persons or institutions are responsible for the project coordination?
1.3.2. Project partner
1.3.3. Does your institution have rules or guidelines for the handling of research data? If yes, please briefly outline them and refer to more detailed sources of information if necessary. Please also indicate, if the rules / guidelines are mandatory or optional.
1.3.4. Who is/are the contact person(s) for data management questions?
1.4. Funding
1.4.1. Who is funding the project?
1.4.2. Is the project within a special funding programme?
1.4.3. Does the funder have rules or recommendations for data management? If yes, please briefly outline them and refer to more detailed sources of information if necessary. Please also indicate, if the rules / guidelines are mandatory or optional.
1.5. Other requirements
1.5.1. Are there requirements regarding the data management from other parties (e.g. the scholarly/scientific community)?
1.5.2. Which are these additional requirements regarding data management?

2. Content classification
2.1. Data
2.1.1. What kind of dataset is it?
2.1.2. Is the dataset being created or re-used?
2.1.3. If re-used, who created the dataset?
2.1.4. If re-used, under which address, PID or URL can the dataset be found?
2.1.5. Which individuals, groups or institutions could be interested in re-using this dataset? What are possible scenarios?
2.1.6. Is the dataset reproducible in the sense that it could be created / collected anew in case it got lost?

3. Technical classification
3.1. Data
3.1.1. When does data collection or creation start?
3.1.2. When does data collection or creation end?
3.1.3. When does data cleansing / data preparation start?
3.1.4. When does data cleansing / data preparation end?
3.1.5. When does data analysis start?
3.1.6. When does data analysis end?
3.1.7. What is the actual or expected size of the dataset?
3.1.8. How much data is produced per year?
3.1.9. Which file formats are used?
3.1.10. Which tools, software, technologies or processes are used to generate or collect the data?
3.1.11. Which software, processes or technologies are necessary to use the data?
3.1.12. Is documentation about relevant software needed to use the data?
3.1.13. Are different versions of the dataset created?
3.1.14. Which versioning strategy is applied for this dataset?

3.1.15. Which technology or tool is used for versioning?

4. Data usage
4.1. Usage scenarios
4.1.1. How / for what purpose will this dataset be used during the project?
4.1.2. How often will this dataset be used?
4.1.3. To what extent will infrastructure resources be required (e.g. CPU hours, bandwidth, storage space... etc.).
4.1.4. Are there actual or potential usage scenarios that could benefit from support by a data management or IT expert, or that even require such support?
4.2. Data storage and security
4.2.1. Where is the dataset stored during the project?
4.2.2. Under which URL can the dataset be accessed during the project?
4.2.3. Are there internal project guidelines for a consistent organisation of the data? If so, where they are documented?
4.2.4. Is there a internal project guideline for naming the data? If so, please briefly outline the naming conventions and, if necessary, link to the documentation.
4.2.5. Who is allowed to access the dataset?
4.2.6. How and how often will backups of the data be created?
4.2.7. Who is responsible for the backups?
4.2.8. Which measures or provisions are in place to ensure data security (e.g. protection against unauthorized access, data recovery, transfer of sensitive data)?
4.3. Data sharing and re-use
4.3.1. Is this dataset interoperable, i.e. allowing data exchange and re-use between researchers, institutions, organisations, countries etc.?
4.3.2. Will this dataset be published or shared?
4.3.3. If no, please explain why not. Please differentiate between legal and contractual reasons and voluntary restrictions.
4.3.4. If yes, under which terms of use or license will the dataset be published or shared?
4.3.5. If there are any restrictions on the re-use of this dataset, please explain why.
4.3.6. When will the data be published (if they are)?
4.4. Collaborative work
4.4.1. Will the data be collaboratively used?
4.4.2. Which platform / tools is / are used for collaboratively working on data and publications?
4.4.3. How is the collaborative work on the same files organised?
4.5. Quality assurance
4.5.1. Which measures of quality assurance are taken for this dataset?
4.5.2. Is the integration between the re-used and newly created data ensured? If yes, by which means?
4.6. Costs
4.6.1. What are the personnel costs for data management associated with the creation or acquisition of data in the project?
4.6.2. What is the amount of non-personnel-costs for data management associated with the creation or acquisiton of data in the project?
4.6.3. What are the personnel costs for data management associated with the the usage of data in the project?
4.6.4. What is the amount of non-personnel-costs for data management associated with the usage of data in the project?
4.6.5. What are the personnel costs associated with data storage and data security in the project?
4.6.6. What is the amount of non-personnel costs associated with the storage of the data sets during the project?

5. Metadata and referencing

5.1. Metadata

5.1.1. Which information is necessary for other parties to understand the data (that is, to understand their collection or creation, analysis, and research results obtained on its basis) and to re-use it?

5.1.2. Which standards, ontologies, classifications etc. are used to describe the data and context information?

5.1.3. Which metadata are collected automatically?

5.1.4. In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

5.1.5. Which metadata are collected semi-automatically?

5.1.6. Which metadata are collected manually?

5.1.7. Are metadata and context information being checked for correctness and completeness?

5.1.8. Who is responsible for documenting the metadata and context information and for checking if they are correct and complete?

5.1.9. What are the personnel costs associated with the the creation of metadata and context information in the project?

5.1.10. What is the amount of non-personnel-costs associated with the creation of metadata and context information in the project?

5.2. Structure, granularity, and referencing

5.2.1. What is the structure of the data? How are the individual components of the dataset related to each other? How is the dataset related to other datasets used in the project?

5.2.2. Will persistent identifiers (PIDs) be used for this data set?

5.2.3. Which system of persistent identifiers shall be used?

5.2.4. Which (sub-) entities / sub units should be referenced using identifiers? Which of those identifiers should be persistent and citable?

5.2.5. Who is responsible for the maintenance of the PIDs and the object maintenance (i.e. who is responsible notifying the PID-Service about object relocation and the new address)?

5.2.6. What are the personnel costs associated with of persistent identifiers in the project?

5.2.7. What is the amount of non-personnel-costs associated with persistent identifiers in the project?

6. Legal and ethics

6.1. General legal issues

6.1.1. Does the legal situation of different countries have to be considered?

6.2. Sensitive data

6.2.1. Does this dataset contain personal data?

6.2.2. Which law applies with respect to the aspects of data protection in the project?

6.2.3. Does the dataset contain "information on racial and ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, health or sex life" (BDSG §3, Abs.9)?

6.2.4. Will the data be anonymised or pseudonymised?

6.2.5. To what extent is the "informed consent" obtained from the persons concerned?

6.2.6. If no "informed consent" is obtained, please give the reasons for not doing so.

6.2.7. Where and how is the "informed consent" documented?

6.2.8. By when will the (unanonymised or unpseudonymised) original data be safely deleted?

6.2.9. Does this dataset contain sensitive data other than personal data?

6.2.10. If yes, please describe the non-personal sensitive data used in the project.

6.2.11. What are the personnel costs associated with the anonymization of sensitive data in the project?

6.2.12. What is the amount of non-personnel-costs associated with the anonymization of sensitive data in the project?

6.2.13. What are the personnel costs associated with other (non-technical) security measures for sensitive data in the project?

6.2.14. What is the amount of non-personnel-costs for other (non-technical) security measures for sensitive data for the project?
6.2.15. Has the project been approved by a research ethics committee?
6.2.16. Is a statutatory approval / permit needed for the research?
6.2.17. If yes, which permit?
6.2.18. If yes, which is the responsible agency?
6.2.19. Is a data access committee needed to handle access requests to the published data of the project?
6.3. Intellectual property rights
6.3.1 Does the project use and/or produce data that is protected by intellectual or does copyright law apply to this dataset?
6.3.2 Do other intellectual property rights apply to this dataset?
6.3.3 Was investigated who the rights owner is?
6.3.4 What are the personnel costs associated with intellectual property rights in the project?
6.3.5 What is the amount of non-personnel-costs regarding intellectual property rights in the project?

7. Storage and long-term preservation
7.1. Selection
7.1.1. What are the criteria / rules for the selection of the data to be archived (after the end of the project)?
7.1.2. Who selects the data to be archived?
7.2. Long-term preservation
7.2.1. Does this dataset have to preserved for the long-term?
7.2.2. What are the reasons this dataset has to be preserved for the long-term?
7.2.3. How long will the data be stored?
7.2.4. How long is it intended that the data remains re-usable.
7.2.5. Where will the data (including metadata, documentation and, if applicable, relevant code) be stored or archived after the end of the project?
7.2.6. Is the repository or data centre chosen certified (e.g. Data Seal of Approval, nestor Seal or ISO 16363)? (If the dataset is archived at several places, you may answer this question with yes, if this applies to at least one of these.)
7.2.7. Have you explored appropriate arrangements with the identified repository?
7.2.8. Shall there be an embargo period before the data are made available?
7.2.9. How will the identity of the person accessing the data will be ascertained?
7.2.10. By when will the data be archived?
7.2.11. What are the personnel costs associated with long-term preservation for the project?
7.2.12. How will the datamanagement costs of the project be covered?
7.2.13. What is the amount of non-personnel-costs regarding long-term preservation for the project?