**GSI Data Management Plan Template: Example Case 1\_2, mid-project**

**Example Info**

*v.2.0 –Case 1\_2 example – Case parameters:*

|  |  |  |  |
| --- | --- | --- | --- |
| *Data Type* | *Data Size* | *Project Status* | *Published data* |
| *Experimental + simulation* | *35TB storage/ 15GB published*  | *After data collection/ During analysis* | *Pre-processed/Result -> Zenodo* |

*Example answers are given in red*

**Preamble:**

A Data Management Plan(DMP) is a comprehensive document that outlines how data are to be handled both during a research project and after the project is completed. This includes how research data will be collected, processed, stored, and shared during and after a research project. The plan typically includes information on the types of data to be collected, data storage and backup strategies, data security and access protocols, and plans for sharing and preserving the data. The goal of a data management plan is to ensure that the data is well-organised, properly documented, and accessible to the research team and other authorised users, while also protecting the privacy and confidentiality of any sensitive information.

The document can be updated throughout the project, and completed by one or more parties.

*Further information can be found on the on the* [*GSI Open Science Webpage*](https://www.gsi.de/work/forschung/open-science)*, and in the* [*GSI Instructions for Data/Software Uploads*](https://doi.org/10.5281/zenodo.7664633)*.*

For comments, questions, and support please contact the Research Data Management Team open-science@gsi.de

*Example answers are given in red*

1. ***General Project Information***
	1. Data Management Plan Version: v01
	2. Principle Investigator: *Andrew Kishor Mistry*
	3. Project Name: *Test of the GSI Data Management Plan*
	4. Do you already have a DMP for this project at this stage? *no*
		1. If yes, please attach the document here
	5. Project Stage *(e.g. Beginning, after data collection, after analysis…)*

*After data collection/During analysis*

* 1. Date: *24.02.24*
	2. Project ID *(e.g. Proposal number)*: *G-22-00XYZ*
	3. Grant Number(s) *(optional)*: *DFG- XYZ*
	4. Responsible GSI department: *RED*
	5. Responsibilities during the course of the project (i.e. who is responsible for research data management in the course of the project)? *PI (Andrew Mistry)*
	6. Responsibilities after project completion

 *Foreseen PI (Andrew Mistry); Otherwise responsibilities will be transferred to co-proposer*

* 1. Is there a collaboration based data management policy for the project? If so please give details. *No collaboration based DMP*
	2. Briefly describe how the data in the research project will be documented during planning, collection, processing, and evaluation*.*

*During the planning phase, a DMP was prepared. After data collection and during analysis, the DMP is updated to reflect changes and evaluate the status of data in the project. Electronic logbooks are being/were used to document all phases of the data collection and analysis stages. The data is backed up in a suitable storage location and this is noted here. A run sheet listing all files and certain metadata pertaining to analysis is available. Weekly meetings with slide decks are held by collaborators analysing the data during the processing stage. Plans on publication are now underway.*

1. ***Short Description of the Project (abstract could be copied here)***

*This experiment will aim at measuring XYZ using the GSI Detector setup coupled to the GSI accelerator. Simulations will be performed to justify and validate results.*

1. ***Data Set Description***
	1. What types of data will be collected or created (e.g. experimental, simulation)? *Experimental and simulation datasets*
	2. Which format(s) will the data have? (e.g. .lmd, .root,. ascii., .png etc*)*

*Experimental has .lmd as raw data format. The pre-processed experimental and simulation data is in .ROOT format. The final plots will be as .svg files.*

* 1. How large is the data set expected to be total raw, pre-processed, and result data?

*We collected 17TB for the experimental data and anticipate 2TB for the raw simulation data (ongoing). The addition processing step for the experimental data required 16TB. In total we require ~35TB.*

* 1. Do you plan to store the data at GSI? *Yes*
		1. If so, specify the storage location and/or medium

*The data is stored on the lustre storage system and the tape robot*

* + 1. How much space will be required?

*17TB is currently stored on the tape robot and 35TB on lustre. We anticipate a further 2 TB for each storage medium will be required.*

* 1. How long will you retain the data? (The minimum requirement at GSI/FAIR is 10 years)

*Data will be stored for 10+ years*

* 1. How will the data eventually be deleted?

*The data will be deleted only if there is a need to free up disk space after the 10 year time period has elapsed.*

1. ***Publishing data***
	1. Will the data be deposited in a data repository? *yes*
		1. If yes which repository will you use? *(e.g. Zenodo, Discipline specific, etc.)*

*Zenodo*

* + 1. If no; please explain here
	1. What format of data do you plan to publish? *(e.g. raw, pre-processed, result data…)*

 *Pre-processed .ROOT files will be published which will contain ROOT trees from the compressed data. In addition, the data from plots will be published in .ascii format as result data.*

* 1. What data size do you expect to publish? *~15GB*
	2. When will the data be published and made available*?*

*At the same time as the final journal article being published*

* 1. Will the data be made Open Access? Under which licence? GSI strongly encourages researchers to publish data under Attribution 4.0 international (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/> If you will publish under any other open access licence, please give the license and the reason. If you cannot publish open access for other reasons, please answer N/A (further questions will be given below).

*The data will be openly available under CC BY 4.0 as recommended*

* 1. Will there be an embargo period for the data?

*Yes, the data can be uploaded to the repository, and restricted access granted to the collaboration and journal reviewers.*

* + 1. If yes for how long for? *Until the article is published*
	1. In addition to the GSI publication repository, will the data be linked to any other external repository? Please specify.

*The data in Zenodo may be linked to the repositories of other collaboration institutions. However, the published data will remain in Zenodo and not be published elsewhere*.

1. ***Findability of the data (Please see*** <https://www.dcc.ac.uk/guidance/standards/metadata> ***for a description of metadata)***
	1. Are there any methods for organising, labelling or describing your research?

*Yes, we have an internal collaboration structure for filenames, formats and description of the parameters.*

* 1. Will you use a known metadata standard or create one?

*We used a local schema for general data information. At the moment, no external schema exists for our data. Therefore, we used our internal collaboration structure and are currently working with external partners to agree upon a common structure.*

* 1. Where will the metadata be documented and stored*?* (The RDM Team can be contacted for assistance if needed)

*The metadata is contained in the logbook for each file. A common file naming conversion is used and a description of the reaction type, file duration and other experimental parameters is given. A summary runsheet is prepared*

* 1. Which type persistent identifier will the published Research Data have? *(e.g. DOI)*

*The published data will use the DOI generated by Zenodo*

1. ***Security***
	1. Are there any security issues relating to the storage of the data? *None*
	2. Who else will have access to the data during this project?

*Only members of the collaboration and if needed external reviewers*

* 1. Will the data be stored elsewhere (pre-publication)?

*The data is being held by external collaboration members for analysis purposes.*

* 1. If so, please specify where and how.

*PhD students and postdoctoral staff external to GSI have taken copies of the data and are performing analysis on them. The data is stored on hard drives and also on the external institutes local file system. The data can only be accessed by members of the collaboration.*

1. ***Interoperability of the data***
	1. Will you use additional internal non-published data?

 *Experiment G-22-XYZ used a similar experimental reaction to ours, with a different experimental setup. We are working with the collaborators from that experiment on the analysis stages.*

* 1. Will you use additional published data from GSI/FAIR?

*We will use a combination of datasets with data collected for this study with a portion of the data collected from GSI experiment SXYZ from 2014.*

* 1. Will you use additional published data external to GSI/FAIR?

*Data collected in external laboratories and published in external repositories may be used in for a comparative study: as a validity cross check of results and to check the consistency of findings. This will be done once our analysis phase is completed.*

* 1. Do you envisage that external collaborations will make use of your data?

*We have had interest from external companies in analysing these datasets with machine learning algorithms to improve the precision of results.*

1. ***Reusability of the data***
	1. Will you provide any support for reuse?

*We will provide contact details to support reasonable questions on reuse. The amount of time and resources spent will depend upon the constraints of our researchers.*

* 1. Will you provide additional software or code for further analysis of the published data?

*The software code needed to analyse the published .ROOT trees will be made available on the GitLab repository, and published to Zenodo with a DOI. This will require CERN ROOT installations from the end user.*

* 1. Will you provide documentation for the reuse of the data?

*We will provide a short list of steps needed to run the .ROOT trees. More detailed documentation will be provided to give a description of the trees structure, and necessary correlation conditions. A link to the CERN ROOT webpage will be provided*.

* 1. If so, how long will you offer support for reuse*? For up to ten years, longer if needed.*
	2. Who will be able to reuse the data?

*The data may be of use to researchers in future work at GSI. This could be aiding in future studies, teaching applications, or demonstration proof of principle purposes. In addition, since we plan to publish pre-processed data and code, which will allow conditions on the data to be modified, we hope to offer a dataset that will be useful to external collaborations with similar research interests, and to give the general public a flavour of our research. The publishing of the result data from plots may also allow plot replication and combination of datasets.*

* 1. GSI strongly encourages researchers to publish data under Attribution 4.0 international (CC BY 4.0) <https://creativecommons.org/licenses/by/4.0/> If you will publish under any other open access licence, please give the license and the reason. If you cannot publish open access for other reasons, please answer N/A (further questions will be given below).

*We plan publish under CC BY 4.0 as recommended*

1. ***Ethics***
	1. Are there any are ethical or legal issues that could have an impact on data sharing? If yes, please specify. *None*
	2. Are there any copyright issues to the data created in the project? *None*
	3. Will you collect and store personal or sensitive data as defined under the terms of the GDPR? (e.g. includes email addresses, phone numbers*)*

*Names and any entries containing personal information in electronic logbooks will be published only with the permission of the persons involved, or redacted.*

* 1. Are there any dual use/technology transfer aspects that mean the data cannot be published?

*We are in discussion with GSI technology transfer office over commercial use of the data after publication*

If yes, and not already done, please contact the technology transfer office, and/or legal department to discuss the requirements.

1. ***Associated Costs***
	1. How much personnel power do you envisage to make your data public? (in person months) *0.5 FTE*
2. ***Additional Comments and Notes***

*This example is a representative example of some possible answers and by no means an exact replication of any given project. Further information and support for completing the DMP can be found by contacting* *open-science@gsi.de*