# *liv*MatS Data Management Plan Template — dtool

This template for a data management plan (DMP) template implements the *livMatS RDM policy* by the dtool research data management ecosystem.

* *Italic entries provide suggestions and are to be adopted per-project.*
* Empty fields are to be filled out with project-specific information.
* Other default text defines unnegotiable content according to the *livMatS RDM Policy*.

|  |  |  |
| --- | --- | --- |
|  | General Information | |
|  | Administrative information such as name of applicant, project number, funding programme, and version of DMP  This DMP implements the *liv*MatS research data management (RDM) policy [1]. | |
| 1 | Data description and collection or re-use of existing data | |
| 1a | How will new data be collected or produced and/or how will existing data be re-used? | * Collaborators are required to package any created data in dtool datasets [2]. They are encouraged to aim for fine granularity (e.g. one measurement -> one dataset). Access to and support for the dtool research data management framework and underlying storage infrastructure bwSFS [3, 4] are provided by livMatS and the University of Freiburg’s computing center. * A dtool dataset is identified by a universally unique identifier (UUID) [5] assigned at creation. * Collaborators are required to track data provenance by documenting the relation of source datasets and derived datasets. The preferred method is listing the persistent identifiers (i.e. UUIDs, see below) of all source datasets within the metadata of a derived dataset under the key ‘derived\_from’. |
| 1b | What data (e.g. the kind, formats, and volumes), will be collected or produced? |  |
| 2 | Documentation and data quality | |
| 2a | What metadata and documentation (e.g. the methodology of data collection and way of organizing data) will accompany the data? | * *Administrative and bibliographical metadata are provided with vocabulary adhering to the DataCite Metadata Schema / DCMI metadata terms (Dublin Core) / [schema.org type Dataset](https://schema.org/Dataset" \o "https://schema.org/Dataset) / other.* * *Descriptive metadata are provided according to discipline-specific ontology XYZ.* * Data and metadata are bundled in self-descriptive datasets at data creation. For this purpose, we use the data management framework dtool. The structure of data and metadata in a dtool dataset is described by Olsson and Hartley [2]. |
| 2b | What data quality control measures will be used? | * A dtool dataset is identified by a universally unique identifier (UUID). The dataset is created, filled, and “frozen”. The latter makes it immutable. Modifications to frozen data are only possible by deriving a new dataset with another UUID from the original data. Data provenance is tracked by indicating the source dataset within the derived dataset’s metadata. |
| 3 | Storage and backup during the research process | |
| 3a | How will data and metadata be stored and backed up during the research? | * Frozen dtool datasets are deposited on the *liv*MatS-wide S3 object storage repository provided on bwSFS via the University of Freiburg’s central computing services. * The bwSFS system holds data redundantly at two different physical locations within the University of Freiburg’s campus. It offers the additional option to georedundantly replicate data to the University of Tübingen. |
| 3b | How will data security and protection of sensitive data be taken care of during the research? |  |
| 4 | Legal and ethical requirements, codes of conduct | |
| 4a | If personal data are processed, how will compliance with legislation on personal data and security be ensured? | not applicable |
| 4b | How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable? |  |
| 4c | What ethical issues and codes of conduct are there, and how will they be taken into account? |  |
| 5 | Data sharing and long-term preservation | |
| 5a | How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons? | * All data are deposited on the *liv*MatS-wide S3 object storage repository provided on bwSFS via the University of Freiburg’s central computing services at [https://s3.bwsfs.uni-freiburg.de](https://s3.bwsfs.uni-freiburg.de" \o "https://s3.bwsfs.uni-freiburg.de) within bucket ‘frct-livmats’. Per default, non-sensitive data on this S3 repository is searchable and readable *liv*MatS-wide. * Datasets on the *liv*MatS-wide repository are ingested by the dtool lookup server [https://livmats-data.vm.uni-freiburg.de](https://livmats-data.vm.uni-freiburg.de" \o "https://livmats-data.vm.uni-freiburg.de:5000) and thus made searchable within the dtool RDM framework. * *Along with textual publications produced in the course of this project, all underlying datasets of raw and derived data are published on open-access platforms like zenodo.org*. |
| 5b | How will data for preservation be selected, and where will data be preserved in the long term (e.g. data repository or archive)? | * Data will reside on the *liv*MatS-wide repository [https://s3.bwsfs.uni-freiburg.de](https://s3.bwsfs.uni-freiburg.de" \o "https://s3.bwsfs.uni-freiburg.de) within bucket ‘frct-livmats’ for at least 10 years. |
| 5c | What methods or software tools are needed to access and use data? | * Datasets are managed via the dtool research data management framework. Olsson and Hartley [2] describe the structure of a dtool dataset on a conventional, hierarchical file system. The dataset bundles data and metadata in a self-descriptive manner and is designed to outlive the particular software used for creation. * All data are held on a software-independent object storage repository (see 5a). The repository is searchable via the lookup mechanisms provided via dtool. Additionally, all data are accessible via the low-level S3 protocol. |
| 5d | How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured? | * All datasets are assigned a Universally Unique IDentifier at creation. * *Published datasets are assigned persistent identifiers according to the particular repository platform’s standards (usually a DOI).* |
| 6 | Data management responsibilities and resources | |
| 6a | Who (e.g. role, position, and institution) will be responsible for data management (i.e. the data steward)? | * The data creator (researcher) is responsible for data quality, dataset creation, dataset freezing, documentation via assigning meaningful metadata, and dataset deposition on the *liv*MatS-wide repository. The *liv*MatS data steward supports the researcher in these aspects. * The central computing center (Rechenzentrum) provides access to bwSFS storage infrastructure. bwSFS offers (geo-)redundancy. * *liv*MatS provides access to the *liv*MatS-wide bwSFS-backed data repository via the dtool RDM framework. * The *liv*MatS data steward assures the availability of above services at all times and supports the researcher with advice, documentation and trainings. * The PI is responsible for implementing, reviewing and revising this DMP. The *liv*MatS data steward may help with this task. |
| 6b | What resources (e.g. financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)? | * The *liv*MatS data steward advises researchers on all data management-related topics such as data creation, documentation, storage, curation, publication. |

# References:

1. [https://www.livmats.uni-freiburg.de/rdm](https://www.livmats.uni-freiburg.de/rdm"\o "https://www.livmats.uni-freiburg.de/en/members/research-data-management)
2. T. S. G. Olsson and M. Hartley, “Lightweight data management with dtool,” PeerJ, vol. 7, p. e6562, Mar. 2019, doi: 10.7717/peerj. 6562.
3. <https://rdmg.uni-freiburg.de/posts/2021/06/bwsfs-storage-for-science/,> accessed 2022-03-04
4. bwSFS, Baden-Württemberg’s Storage for Science, funded by the Deutsche Forschungsgemeinschaft (DFG) under project number 405998531.
5. P. Leach, M. Mealling, and R. Salz, “A universally unique identifier (uuid) urn namespace,” RFC 4122, Jul. 2005. doi: 10.17487/RFC4122. [Online]. Available: [https://www.rfc-editor.org/info/rfc4122]("https://www.rfc-).