Deliverable D6.1

Data Management Plan

Deliverable

Number and title:	D6.1 – Data Management Plan		
Work package:	WP6 (Project Management)		
Editor:	Arnd Hartmanns (UT)		
Due date (GA):	M6 (2022-03-31)	*Dissemination level:*	Public
Due date (revised):	2022-10-14	*Version:*	1.0
Submission date:	2022-09-08	*Pages:*	9

Version history

Version	*Date*	*Notes*
1.0	2022-09-08	First official release

Project

Title:	Models in Space Systems: Integration, Operation, and Networking		
Acronym:	MISSION	*Start date:*	01-10-2021
GA no.:	101008233	*Duration:*	48 months
Call:	H2020-MSCA-RISE-2020	*Website:*	mission-project.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

MISSION. Designing reliable spacecraft involves accounting for a wide variety of faults and environmental challenges, such as radiation and space debris. To overcome these obstacles, spacecraft are often over-engineered, significantly adding to cost and complexity. Increasingly, the commercialisation of space now involves small businesses using existing solutions to launch constellations of small satellites. This approach can reduce cost, weight, size and production time, but it also presents challenges that increase uncertainty and reduce reliability. The MISSION project develops the technologies necessary to assist in critical aspects of spacecraft design. The goal is to improve reliability and effectiveness for future space missions of the European and global space industry.

This deliverable. The Data Management Plan (DMP) defines the data management system and presents the "life cycle" of data generation, collection, processing, preservation, sharing and reusing. The DMP complies with the FAIR principles for data management. It describes the handling of research data during and after the end of the project; what data will be generated, collected, and processed; which methodology and standards will be applied; whether data will be shared/made open access; and how data will be curated and preserved also after the end of the project. This DMP applies primarily to the data needed to validate the results presented in scientific publications resulting from the project. The current plan will ensure that all publications are open access and that accompanying datasets will by default be made available.

This document. This report follows the template provided by the European Commission. During the project, it will be updated in line with the periodic reporting. Section 1 describes the details of the data types for all the flagship codes. We describe the implementation of the FAIR principles in Section 2. In Section 3, the allocation of resources is defined, while the data security is described in Section 4. Section 5 addresses the ethical aspects, while other items, such as local data management procedures, are mentioned in Section 6. Section 7 concludes the document.

1 Data Summary

Data collection and generation has an important role in the MISSION project. Table 1 provides an overview of the data that is expected to be generated, collected and utilized throughout the project, along with some information regarding their origin, form, format, software used, size and storage and backup locations. The collected data is not expected to include personal data in terms of the European General Data Protection Regulation (GDPR). In addition the the information in Table 1, we provide further background on the different types of data below.

Models and model templates. The MISSION project is centred around model-based methods and the creation of models and model template libraries for the New Space domain. Therefore, many tasks of the project implicitly or explicitly (especially in WP3) generate models. Task T4.7 in particular is to create a library of "models-off-the-shelf" (MOTS). These models are useful firstly for the project partners for the implementation of the respective tasks, and later—especially in the case of the MOTS library—for external users of the project's results such as third companies implementing MISSION's MOTS-based paradigm.

Software tools and libraries. To support the MOTS approach, the MISSION project's tasks include the development of software tools (such as extensions of the Modest Toolset in tasks T2.6 and T3.4, and of ePMC in T4.5 and T4.6) and libraries (e.g. as extensions to the ROLL library in T1.5). These implement newly developed algorithms and provide reference implementations for the modelling formalisms proposed by MISSION. They will again be useful for all project partners during the implementation of the project, and ultimately to both other researchers who will build upon MISSION's (more theoretical) scientific results as well as to other companies who implement the (practical) MISSION approach for their own (space) missions. We aim for open-source releases of the software developed in MISSION, but

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

Type of data and related materials	Forms	File format	Software tools	Size	Storage and backup
Models and model templates	Model files	Text files, XML, JSON, etc.	Text editors	1-100 GB	Beneficiaries' networks/long-term archives
Software tools and libraries (source code)	Source code	Text files	Text editors, compilers and build systems	10-100 GB	Beneficiaries' networks/long-term archives
Software tools and libraries (binaries)	Binary files	Binary executables, shared and static libraries	n/a	10-100 GB	Beneficiaries' networks/long-term archives
Scientific publication drafts/sources	Text	Text files (LaTeX format), Microsoft Word	Text editors, LaTeX, Microsoft Word	$\leq$ 1 GB	Beneficiaries' networks
Final scientific publications	Text	PDF	PDF	$\leq$ 1 GB	Publishers' systems, institutional and EU repositories
Experimental data related to scientific publications	Text, tabular data, scripts, and others (specific to publication)	Text files, CSV format, Microsoft Excel, and others	Text editors, Microsoft Excel, and others	100 GB-2 TB	Beneficiaries' networks/long-term archives

Table 1: Outlook of expected types of data within the MISSION project

recognise that this will not always be possible, especially where existing projects (such as the Modest Toolset, in which copyright is shared by multiple institutions and which is currently closed source with source access given to research partners on demand, such as those within the MISSION project as per the consortium agreement) and intellectual property of the industrial partners are concerned.

Experimental data. Many MISSION tasks, and consequently the resulting publications, are based on performing experiments to evaluate and validate the technology (such as models, algorithms, and software tools) developed in the project. Examples include the empirical evaluation of fault tree analysis methods in T2.5 and the measurements for CGR planned in T3.2. We collect this data for the purpose of performing the experiments and for being able to repeat our experiments ourselves, and we publish and archive this data to allow third parties, in particular other researchers, to reproduce our results for validation, and to build further research upon our results and our experimental setup. Typically, the whole experimental setup—all scripts, input files, and our original outputs—related to a submitted research paper will be packaged into an "artifact" that is often evaluated in an artifact evaluation as part of the reviewing process, and that will be long-term archived with a DOI in a repository like 4TU.ResearchData, Figshare, or Zenodo. For more details, we refer the reader to the ACM's overview of artifact review and badging at acm.org/publications/policies/artifact-review-and-badging-current.

Scientific publications. Conference and journal publications are the main scientific results of MISSION. They are obviously useful for other researchers in MISSION's focus area and the formal methods community in computer science, and to industrial users seeking insights into the MISSION approach.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

Re-use of existing data. MISSION will reuse existing data in several tasks. Some of this data is present at the project's industrial partners, but not currently publicly available (in particular since it often contains confidential operational data); this data will be processed by the MISSION partners in several tasks to be transformed into new data that will be openly published as part of the types of data described above. An example is the mission and topology data collected in T3.1 that has already been openly disseminated as part of deliverable D3.1. In terms of concrete publicly available data, we currently expect the following reuse as part of the project:

- Models from the Quantitative Verification Benchmark Set: these are text and JSON files available at qcomp.org/benchmarks under a CC-BY license that we will use for benchmarking our algorithms and tools.
- The Modest Toolset, distributed in binary form at modestchecker.net, serves as the base for several implementation tasks, e.g. T2.6. Its source code is stored in a Git repository at USAAR, and access will be granted to project partners as necessary, as guaranteed by the consortium agreement.
- We will use the µD3TN (gitlab.com/d3tn/ud3tn) and aiodtnsim (gitlab.com/d3tn/aiodtnsim) open-source implementations of DTN protocols.
- We might use EPMC (github.com/iscas-tis/ePMC) for the translation to IO-JANI, by implementing a plugin for it, and potentially for other functionalities.
- The ROLL library, licensed under the GPL and available at github.com/iscas-tis/roll-library, will be extended as part of T1.5.
- The FIG Tool, licensed under GPL and available at git.cs.famaf.unc.edu.ar/dsg/fig, will be used and extended in the context of WP2 in e.g. T2.5 and T2.6.
- DtnSim (bitbucket.org/lcd-unc-ar/dtnsim), a discrete-event open-source simulator for DTN protocols, and Orekit (orekit.org), an open-source orbital propagator, will be used in WP3.
- The libraries and code related to RUCoP, released under MIT license and currently available online at bitbucket.org/nandoraverta/bruf-withcopies19/, will be used and extended in the context of WP3 in e.g. T3.2 and T3.3.

2 FAIR Data

This section describes the implementation of the FAIR principles as part of data management in MISSION to make the data associated to our scientific publications findable, accessible, interoperable, and reusable.

2.1 Making Data Findable, Including Provisions for Metadata

For all our publicly archived data, which usually is assigned a DOI, we will use the metadata schemes of the respective repositories which follow accepted standards, e.g. the DataCite Metadata Schema (see schema.datacite.org) in the case of Zenodo (cf. about.zenodo.org/principles). The names of our archived artifacts will be named so that the connection to the corresponding publication or deliverable is immediately visible, by following a naming scheme such as "*Title of Publication (Artifact)*". The repositories we plan to use assign DOIs, and therefore provide clear versioning in case we update an archived dataset.

Where project-internal storage and exchange of data is concerned, we will use clear folder structures that include simple text files describing the data. We will set up a central spreadsheet tracking all of our internal data collections, their locations, and the corresponding partner contacts to make them findable for consortium members. In most cases, we will use Git repositories with their inherent versioning capabilities.

To facilitate the usage and re-use of the stored data, all stored (internal and archived) data will be

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

accompanied by a text-based description of the data detailing its source, format, and how the data can be interacted with.

2.2 Making Data Openly Accessible

We strive for open access to our data: MISSION will deposit and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate—free of charge for any user—data resulting as well as results from the funded project. All publications resulting from the project will be open access (at least on the "green" level) as required by the Grant Agreement. For the data types of models, publications, and experimental data, the default license to be used in MISSION is the CC-BY 4.0 Creative Commons license (see creativecommons.org/licenses/by/4.0). For software, we plan to use open source licenses approved by the Open Source Initiative (OSI) such as the GPL, MIT, or Apache licenses; see opensource.org/licenses for details. A dataset's license will be included in the dataset itself in an obvious location such as a file named `License.txt` at the root of the folder structure.

The main limitations to making our data open access concern experimental data collected by the industrial partners, and the source code of some software tools and libraries. For the latter, we refer to the corresponding paragraph in Section 1. For the former, if publication of the full data is not possible to protect the interests of a partner, we will strive to nevertheless make a redacted dataset available. In particular, we will always make enough data available for third parties to be able to reproduce a sufficient subset of our experiments to confirm the findings we present in our publications.

A list of exceptions to the open access rule of MISSION will be added here in future revisions of this document.

2.3 Making Data Interoperable

Data will be stored in formats as outlined in Table 2 to allow the re-use of any data as appropriate. Most of these formats are simple text-based formats (such as models in the JSON-based JANI format, see jani-spec.org, or CSV files for tabular data), supported by open-source software. We aim to avoid reliance on formats requiring additional software such as Microsoft Excel, noting however that today's standard file formats of Word and Excel are also standardised and XML-based.

2.4 Increase Data Re-Use

All our data will be clearly and by default openly licensed as described in Section 2.2. We do not plan to make use of embargoes on our data and publications. Usability by third parties, also after the end of the project, will not be restricted, with the exceptions outlined in Section 2.2 concerning the open publication of some data in the first place. The duration of storage of our archived data depends on the policies of the repository used; in the case of the 4TU.ResearchData repository, for example, a duration of at least 15 years is guaranteed (see data.4tu.nl). The quality of our data will be assured internally, by discussing and assigning reviews on the state of our data at meetings of the project's general assembly, and externally by artifact evaluation processes for artifacts accompanying conference publications.

3 Allocation of Resources

Every partner is in principle responsible for organising the appropriate storage of the project's internal data managed by that partner, including the costs for hardware, maintenance, administration, and backups. All partners are aware that the project's institutional funds can be used to fund the costs of making data related to the MISSION project FAIR as described above, in particular including the costs of making publications open access. Several of the partners are already able to publish open access for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

free in most of the relevant journals by institutional agreements (such as the ones between the Dutch universities and the ACM and Springer), and many conferences that the participants typically publish at (such as CAV and all ETAPS conferences) are already gold open access without extra fees.

In the following, we summarise the resource allocation plans per project partner.

UT. Internal data will be stored on the UT network at the UT data centre, located at the UT itself and ISO 27001/NEN 7510 certified. Storage time for backups is 28 days. Where extra storage is needed (which we do not expect for MISSION), the costs are between around 140 €/TB/year (normal quality) and 600 €/TB/year (high quality). UT researchers can archive datasets in the 4TU.ResearchData repository at no charge up to 1 TB/year.

RWTH. MISSION will be supported by the Research Data Management team at RWTH, see rwth-aachen.de/Forschungsdatenmanagement.

USAAR. The hosting of the project website at mission-project.eu, where publications are listed as part of the mission-project.eu/publications page, at USAAR comes with monthly costs on the order of 20 €.

D3TN. Internal data will be stored on D3TN's infrastructure that is locally maintained. Open source code and contributions to open source projects are stored in their corresponding public repositories on platforms such as GitLab and GitHub.

GOM. Data will be stored in local servers.

UNC. Internet data will be stored in local servers at FAMAF which are locally maintained. In addition, publications will be also stored at the UNC Digital Repository following the corresponding Open Knowledge Policy.

URNC. Internet data will be stored in local servers at UNRC which are locally maintained. In addition, research results will be also stored at the UNRC Digital Repository following the corresponding Open Access Policies.

ASC. Internal data will be stored in local servers at Ascentio which are locally maintained.

INVAP. Internal data will be stored in local servers at INVAP which are locally maintained.

IISG. Internal data will be stored in the local network at ISCAS or in public repositories on platforms such as GitHub. For the IISG-affiliated researchers involved in MISSION, the costs for publishing papers are usually covered by ISCAS, instead of IISG, by means of other projects.

4 Data Security

We already described aspects of data security in Section 3 above as they require dedicated resource allocation. Appropriate measures related to virus protection, data encryption, and password-secured files and computers will be implemented at all MISSION partners. In this section, we list the specific provisions on data security especially for long-term archiving per project partner.

UT. The UT research data management policy and faculty data policies demand preservation in a trusted data repository. UT chooses to use the 4TU.ResearchData for beta-sciences or technical-sciences data, and thus as its preferred repository for data arising in the MISSION project where UT researchers are leading the implementation. UT researchers can archive datasets in the 4TU.ResearchData repository at no charge up to 1 TB/year. The repository provides DOIs to all archived datasets for easy retrieval and sustainable linking and citation.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

RWTH. Research results will all be made publicly available via the open-access `arXiv` platform. Papers are archived in the data repository of RWTH's central library publications.rwth-aachen.de. Prototypical software tools and benchmarks will be part of artifacts that accompany our research papers so as to ensure reproducibility of the reported results. This archived data will be linked to publications via the RWTH library's publication repository. Our developed software tools will be open access and are made publicly available.

USAAR. MISSION-related data is stored and archived in Git repositories in a GitLab instance hosted at USAAR's Hochschul-IT-Zentrum (HIZ), which also provides HDD- and tape-based backups available for a minimum of 90 days.

D3TN. Data will be stored in redundant, access-protected local repositories that are locally managed.

GOM. Data will be stored in local repositories.

UNC. Data will be stored in local repositories and the UNC Digital Repository following the University and Faculty policies of data management. Additionally, the university leverages Google-based services (G Suite) to safely and reliably store information.

UNRC. Data will be stored in local repositories and in the UNRC Digital Repository at repodigital.unrc.edu.ar following the university's and faculty's data management policies.

ASC. Software will be stored and maintained in Ascentio's local GitLab repositories. Documentation generated by Ascentio will be stored in the Alfresco DMS (document management system) in local servers. Ascentio's own security rules and policies (internal document `SWM-K-PRO-P-D01-005_v2.0_GestionDeLaConfiguracion`) will be taken into consideration for data management.

INVAP. Data will be stored in local repositories and the INVAP Digital Repository following INVAP's policies of data management.

IISG. At IISG there is no support for long-term data storage and availability; we will probably use external resources for getting this (such as GitHub for code and models and Zenodo for artifacts).

5 Ethical Aspects

The main ethical concern identified for the MISSION project is related to the involvement of third-country partners and the dual-use potential of the MISSION research. This may concern the handling and sharing of our data across borders. However, we do not expect these issues to arise in (most of) our implementation of the project as outlined in our clarifications concerning the pre-grant requirements. These issues are separately tracked in the ethics-related processes and deliverables of MISSION. Below, we summarise any specific ethics-related policies and concerns as pertaining to the individual partners.

UT. Ethical issues and codes of conduct applicable to activities involving UT researchers in MISSION are described in the EEMCS faculty's Research Data Management Policy and its Protocol for Assessing the Ethical Permissibility of Proposed Research.

USAAR. The chair of Holger Hermanns participating in MISSION for USAAR is running the course "Ethics for Nerds", see ethics-for-nerds.de. The faculty runs an Ethics Review Board, see erb.cs.uni-saarland.de, which includes experts who also are participating in the MISSION project. The "Kommission für die Ethik sicherheitsrelevanter Forschung" at UdS, see uni-saarland.de/verwaltung/wissenschaftliche-integritaet/sicherheitsrelevante-forschung.html, has been and is giving ethics advice on space-related research projects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

UNC. Ethical issues and codes of conduct applicable to activities involving UNC researchers in MISSION are described in several UNC regulations. The main documents are OHCS 9/2012 (TO RR-2021-1303-E-UNC-REC) and RHCS 724/2018.

UNRC. Ethical issues and codes of conduct applicable to activities involving UNRC researchers in MISSION are described by the UNRC Research Ethics Committee regulations, see the information available at unrc.edu.ar/unrc/coedi/enlaces.html.

INVAP. INVAP's internal regulations will be applicable to activities involving INVAP researchers in the MISSION project.

6 Other Issues

We finally describe any use that the MISSION partners make of other national, funder, sectorial, or departmental procedures for data management.

UT.

- UT research data management policy: utwente.nl/en/lisa/library/miscellaneous/docs-ru/research-data-policy-ut.pdf
- Each ESR linked to the UT is required to follow a course on data management: utwente.nl/en/ctd/courses/1000227/data-management-bootcamp.
- UT complies with the European Code of Conduct Scientific Integrity: utwente.nl/en/organisation/about/integrity/scientific-integrity/downloads/allea-european-code-of-conduct-for-research-integrity-2017.pdf
- UT complies with the Netherlands Code of Conduct Scientific Integrity: utwente.nl/en/organisation/about/integrity/scientific-integrity/downloads/netherlands-code-of-conduct-for-research-integrity-2018-uk.pdf

RWTH. The research data management strategy of RWTH adheres to the FAIR principle. This concept is embedded in the technical and organisational structure of RWTH Aachen University.

- The latest developments on research data management at the RWTH can be found at rwth-aachen.de/cms/root/Forschung/Forschungsdatenmanagement/~vfgd/Blog

USAAR.

- Principles of good scientific practice at UdS: uni-saarland.de/fileadmin/upload/verwaltung/fundstellen/Forschungsangelegenheiten/DB01-342.pdf

UNC.

- UNC research data management policy: digesto.unc.edu.ar/consejo-superior/honorable-consejo-superior/resolucion/1365_2017
- UNC Code of Conduct: digesto.unc.edu.ar/rectorado/rectorado/resolucion/1303_2021_1
- UNC Copyright Protection Program: digesto.unc.edu.ar/consejo-superior/honorable-consejo-superior/resolucion/724_2018

UNRC.

- Research Data Access and Management Policy: repodigital.unrc.edu.ar/PIAA_v5_UNRC_RCS0202-2021.pdf
- UNRC Research Ethics Committee Regulations: unrc.edu.ar/unrc/coedi/docs/ReglCOEDI.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.

7 Conclusion

This document describes the first setup of the Data Management Plan according to the EU template. It follows the FAIR principles and is a living document that will be updated, if necessary, throughout the entire project. During the project, more information can be added to this document or released via one of the project deliverables.

We also note that the MISSION website was launched on December 20, 2021: Three months into the project, MISSION thus already has an attractive public presence that helps disseminate the results and data produced by the project, and that is now ready to be filled with content as the project progresses.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 101008233.