

Historique du document			
Date	Version	Auteur	Commentaires
12/02/2021	1.0	GT science ouverte	Création du document

Data management plan GUIDE Observatoire de Paris

General information

Describe here the project outline. Additional information may be added depending on the project structure.

Permanent identifiers :

ORCID : see [DOI fiche]

Affiliations : see in the appendix the scientific departments' affiliations approved by the Paris Observatory

see also PSL tool : <https://psl.eu/charte-de-signature-scientifique-de-psl>

ROR identifiers :

Paris Observatory	https://ror.org/029nkcm90
GEPI	https://ror.org/01920cw75
IMCCE	https://ror.org/002zc3t08
LERMA	https://ror.org/01g5pq328
LESIA	https://ror.org/02eptjh02
LUTh	https://ror.org/00bbtde36
SYRTE	https://ror.org/03tdef037
USN	-

Useful links :

<https://ror.org/>

Data description and collection or re-use of existing data

In this section, all scientific research data objects shall be described. This includes any observations, measurement, model run, catalog, etc, derived in the course of the project.

Software libraries developed in the course of the project shall also be described in this section.

Data Collection	Research data objects (what data will be collected or created)	List of data collections (see attachment) Type (code, text...) File format (storage, backup, access) Estimated data volume Content, coverage Versioning Brief description
	How will the data be collected or created	Data production methodologies (experiments, observations...) Structure of folders and files Versioning Quality process
	Instruments used	Camera, spectrometer, telescope... Spacecraft Ground stations Code
	Data description	Datasets naming rules (no special characters or spaces)
	Data property	Contract agreement, legal regimes
	Re-use of existing data?	
	General data policy	Policy of funders or institutions
	Scientific publications	Data papers...

- How will new data be collected or produced and/or how will existing data be re-used?
 - Explain which methodologies or software will be used if new data are collected or produced.
 - State any constraints on re-use of existing data if there are any.
 - Explain how data provenance will be documented and described.

See : **DOI fiche provenance**

- Briefly state the reasons if the re-use of any existing data sources has been considered but discarded
- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and re-use.
- What data (for example the kind, formats, and volumes), will be collected or produced?
 - Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.
 - Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf). Specify the MIME-Type. It is recommended to use community standard file formats (e.g., for astronomy: FITS files).

- Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organization, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.
- Give preference to open and standard formats as they facilitate sharing and long-term re-use of data (several repositories provide lists of such ‘preferred formats’). Include a reference for the format description (file format, data formatting inside file and metadata dictionary).
- Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).

Useful link : <https://voparis-wiki.obspm.fr/display/VES/Data+Formats+and+MIME+Types>

Recommended way of describing data in a DMP: a short textual explanation of your data and how you create or collect it, followed by a table or the like that describes key aspects of the data and the collection process in a structured way (see example in appendix).

Example textual explanations of your data:

“I will use the following data (add columns as needed): “

“The file formats described above are proprietary formats. To ensure that my data will remain accessible and reusable, I will export all data in proprietary formats to [RELEVANT OPEN FORMATS] after the end of the project. I will use existing data, namely [DETAILED DESCRIPTION]. The license of this data is [DESCRIPTION OF LICENSE], meaning that I am allowed to [DESCRIPTION OF WHAT I CAN DO WITH THE DATA]”

“I use personal data, namely [DESCRIPTION]. Throughout this DMP, I will include information on how I plan to safeguard this personal data throughout the research process and after the end of the project”

Documentation and data quality

FAIR metadata

In the frame of the data distribution step, a catalog of the data products shall be provided. It shall include metadata that allows to find the data with scientifically relevant parameters. PADC recommends to follow the ObsCore data model for astronomical observation, the EPN-core data model for solar system science related observations, or the SPASE data model for space physics and space weather, VAMDC, for atomic and molecular physics. More detailed guidance can be obtained from PADC. There may be several metadata catalogs depending on the architecture of the project. This metadata is considered as a research data object in this DMP.

The PADC team can help to define the relevant interoperability standard for this project. Such a metadata catalog is a key part of the interoperability layer of your datasets, enabling data discovery and re-use.

See : DOI fiche FAIR

Useful links :

<https://fairsharing.org/>

http://references.modernisation.gouv.fr/sites/default/files/Referentiel_General_Interoperabilite_V2.pdf

- What metadata will accompany the data?
 - Indicate which metadata will be provided to help others identify and discover the data.
 - Indicate which metadata are included in the data products, and those gathered in catalogues associated to the data collections.
 - Use community metadata standards where these are in place and indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.
 - Consider what other information is needed to enable re-use. This should include provenance information, i.e. what was used and generated to produce the data? following which process? This may also include details on the methods and configuration used to collect and process the data, definitions of variables and parameters, units of measurement, and so on.
 - Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.

- What documentation (for example the methodology of data collection and way of organising data) will accompany the data?

Type of documentation necessary to help the secondary users (author of the data, title, date, access conditions, methodology used, procedural information, file type, codes, abbreviations, version of software...)
- What versioning policy is applied to the data, software and documentation ?

Are all version kept and accessible ? how to link data product versions, software versions and documentation ?
- What data quality control measures will be used?

Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.

Example textual explanations of your documentation & metadata:

“I will add the following documentation to my data:

- *Project-level documentation: a README.txt file contained in [LOCATION], based on the template README.txt provided by research support staff at KU Leuven and adapted to the needs of my project and my discipline.*
- *File-level documentation: additional README.txt files contained in [LOCATIONS]”*

“I will add metadata to my data as follows:

I will add metadata to references using [SOFTWARE OR METHOD].

I will add metadata to image files using [SOFTWARE OR METHOD].

When sharing my data, I will include rich metadata by [METHOD].

I will use a metadata standard appropriate to my discipline, namely [STANDARD] [LINK TO ONLINE DOCUMENTATION].

Since my discipline does not have appropriate metadata standards, I will use [BASIC STANDARD] [LINK TO ONLINE DOCUMENTATION] as a basic standard.

Since my discipline does not have appropriate metadata standards, I will seek the advice of colleagues in my discipline and research support staff at my institution to decide which metadata standard (if any) is appropriate for me”

Storage and backup during the research process

Preliminary discussion must be initiated with the target data repository in the course of the preparation of the project. Discipline repositories are enabling data sharing practices and policies in line with the corresponding science community standard. They thus shall be preferred to generic repositories. The PADC team can help you to select a right repository for your project.

Funding agencies have requirements in terms of data availability. They often require that the data is “as open as possible, as closed as necessary”. Check the call documentation. The same rules usually apply to repositories.

- How will data and metadata be stored and backed up during the research?

Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations:

- Sufficient storage or need to include additional charges (storage on laptops, computer hard drives, external storage...)
- Back up of the data
- How many copies?

Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

- How will data security and protection of sensitive data be taken care during the research?
 - Explain how the data will be recovered in the event of an incident and who will be responsible for backup and recovery
 - Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships
 - Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed
 - Explain which institutional data protection policies are in place.

Data Protection Officer Contact: dpo.obs@obspm.fr

Example textual explanations of your storage & back up:

“During the research, my data will be stored using [STORAGE SOLUTION]. This storage solution is appropriate for my purposes because [INFORMATION ON BACKUPS, COSTS, SIZE OF STORAGE].”

“ My data contains personal information, namely [DESCRIPTION]. I chose a storage solution with the appropriate level of security, based on the recommendations made by my institution/research support staff/... “

“Especially for data containing personal information, I have assured that only authorized persons can access the data by [DESCRIPTION OF SECURITY PROCEDURES]”.

Legal and ethical requirements, code of conduct

- If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

This section may not apply to science data, but other dataset (protected data, nominative data, strategic data...), such as project user personal data, website access statistics, etc, must be carefully considered with the Data Protection Officer (DPO) of Observatoire de Paris, as required by the EU GPDR regulation: dpo.obs@observatoiredeparis.psl.eu

Ensure that when dealing with personal data data protection laws (for example GDPR) are complied with:

- Gain informed consent for preservation and/or sharing of personal data
 - Consider anonymization of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data)
 - Consider pseudonymisation of personal data (the main difference with anonymization is that pseudonymisation is reversible)
 - Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party)
 - Explain whether there is a managed access procedure in place for authorised users of personal data.
- How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

See : DOI fiche [prop intel données](#) et DOI fiche [licences](#)

For any question : srcv.obs@obspm.fr or la.bibliotheque@obspm.fr

Copyright and Intellectual Property Rights issues		Who owns the data? License for use and reuse? Restrictions on the reuse of the data
Data re-use	License	Creative commons, open data commons, specific licenses https://opensource.org/licenses https://creativecommons.org/

- Explain who will be the owner of the data, meaning who will have the rights to control access
- Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and re-use licenses.
 - Creative commons, open data commons, specific licenses
 - <https://opensource.org/licenses> (mostly for software licenses)
 - <https://creativecommons.org/> (mostly for documents and data products)
- Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement
- If applicable, indicate whether intellectual property rights (for example Database Directive, sui generis rights) are affected. If so, explain which and how will they be dealt with
- Indicate whether there are any restrictions on the re-use of third-party data.

- What ethical issues and codes of conduct are there, and how will they be taken into account?
 - Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning
 - Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.

Data sharing and long term preservation

Paris Astronomical Data Centre (PADC) hosts and distributes data at Observatoire de Paris. Data distribution uses interoperability standards originating from the IVOA (International Virtual Observatory Alliance), but also uses dedicated standards from VAMDC, OGC, Europlanet-VESPA, SPASE, etc... PADC ensures long time preservation and sustainability of the data and their metadata. It is composed of the scientific departments of Observatoire de Paris and their science themes. PADC is involved in many European and international projects. The sustainability of PADC is ensured by permanent civil servant (36 researchers and 26 IT engineers, part time) positions at Observatoire de Paris, which has kept its duties and buildings since 1667.

Its IT infrastructure for storage, infrastructure for data distribution and computing facilities are hosted inside the computing centre of Observatoire de Paris. The data are stored and replicated on two physical sites distant by more than 7 km and independently connected to the internet. The access to services is enabled by virtual servers that can be easily reconstructed and that are independent of the physical infrastructure hosting them. Their configuration is managed using Puppet, which allows to automatically install applications, monitoring and backup systems, and ensure straightforward recover plans. The data are stored on a ZFS file system. This solution enforces data integrity. In addition to the two instances of disk storage area, PADC has a system of storage on tapes using a storage virtualization from Active Circle. This system allows PADC to ensure the safety and the preservation of the tapes while keeping access flexibility comparable to disk access thanks to a virtualized file catalog.

Most data repositories are proposing archiving plans, as part of the data hosting and distribution.

- How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Open access	General policy	Will the data be openly available? If not, why Level access (all, members of the project...) Open source code?
	Sharing the data	With whom Under what conditions Sharing data via a repository, handle requests...?
		Restriction for some datasets? When will the data be available (embargo)? Exclusivity use? For how long and why? Sharing agreement required?
	Accessibility of the data and metadata	Choice of a repository or archive : ESA Science Data Archive, NASA Archive, Archive CDP, Archive CDS...
Access procedures	Methods, software tools Description of the conditions for access (machine readable license...) If restrictions on use, how will access be provided	

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalogue, use of a secure data service, direct handling of data requests, or use of another mechanism)
 - Outline the plan for data preservation and give information on how long the data will be retained
 - Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents
 - Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions.
- How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Selection	Which data are of long-term value and should be retained, shared, and/or preserved?	Which data may be reused Which data may be kept, for how long Changing files and formats in order of sharing or preservation Estimation final or annual volume Preservation duration
-----------	---	--

- Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes
 - Indicate how it will be decided what data to keep. Describe the data to be preserved long-term
 - Explain the foreseeable research uses (and/or users) for the data
 - Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.
- What methods or software tools are needed to access and use data?
 - Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data
 - Indicate whether data will be shared via a repository, requests handled directly, or whether another mechanism will be used?
 - How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

For the DOI, ask PADC : padc.doi@obspm.fr

See : DOI fiche DOI

- Explain how the data might be re-used in other contexts. Persistent identifiers should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and re-use
- Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.

Example textual explanations of your data selection, data sharing & preservation:

“After the end of the project, I will store my data on [STORAGE SOLUTION].”

“I will ensure that all relevant data is preserved by selecting data based on [CRITERIA]. I will safely dispose of other data by [METHOD].”

“I will ensure that personal information is properly protected by clearly indicating which files contain personal information in the course of the project and anonymizing all relevant files using [PROGRAM OR METHOD].”

“I will seek the advice of research support staff/IT support staff/legal support staff at my institution to properly anonymize relevant files/choose the correct storage solution.”

“I will share data via the appropriate repositories, namely [REPOSITORY].”

“I will apply the FAIR principles as thoroughly as possible when sharing my data: I will ensure the data is findable by sharing it using a solution that gives the data a DOI.”

“I will ensure the data is accessible by clearly indicating how others can get access to the data.”

“I will ensure the data is interoperable by using open file formats and standards whenever possible.”

“I will ensure the data is reusable by connecting it with an appropriate license, documentation, and metadata.”

“I will not share [PART OF THE DATA] because [COMPELLING REASON].”

“I will share my data under a [NAME OF COPYRIGHT LICENSE] license.”

“I will not share my data openly for [COMPELLING REASON], but I will make the dataset available to others who contact [INDIVIDUAL RESPONSIBLE FOR DATA].”

“I will share the metadata about my data on [ONLINE LOCATION] and clearly indicate how others can obtain the full data.”

Data management responsibilities and resources

- Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

Responsible for data management	<p>Roles and responsibilities for all activities (data capture, metadata production, data quality, storage and backup, data archiving, data sharing...)</p> <p>Who is responsible for implementing the DMP, and ensuring it is reviewed and revised?</p> <p>Responsibilities of the partners in collaborative research projects?</p> <p>Contract agreement between partners?</p>
---------------------------------	--

- Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing. Name responsible individual(s) where possible
 - For collaborative projects, explain the co-ordination of data management responsibilities across partners
 - Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised
 - Consider regular updates of the DMP.
- What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?
 - Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. Carefully consider and justify any resources needed to deliver the data. These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges and also costs of long term preservation (software, hardware, technical expertise...)
 - Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered.

Example textual explanations for responsibilities & resources:

“In the course of the research, XXX [NAME (EMAIL ADDRESS)] is responsible for implementing research data management.”

“XXX, [NAME (EMAIL ADDRESS)], is responsible for ensuring that the principal investigator implements research data management.”

“After the research project ends, preservation of data is the responsibility of XXX.”

“Projected costs for storing data for the required number of years after the project: [NUMBER IN EUROS] The storage costs will be paid from [SOURCE OF FUNDS]”

Pour toute question, vous pouvez envoyer un mail à l’adresse
la.bibliotheque@obspm.fr

Laboratoire	Signature proposée
Galaxies, Etoiles, Physique, Instrumentation	GEPI, Observatoire de Paris, Université PSL, CNRS, 75014 Paris [ou 92190 Meudon], France
Institut de Mécanique Céleste et de Calcul des Ephémérides	IMCCE, Observatoire de Paris, Université PSL, Sorbonne Université, Univ Lille, CNRS, 75014 Paris, France
Laboratoire d'Étude du Rayonnement et de la Matière en Astrophysique et Atmosphères	<p>LERMA, Observatoire de Paris, Université PSL, Cergy Paris Université, Sorbonne Université, CNRS, 75014 Paris [ou 92190 Meudon], France</p> <p>Pour les équipes hébergées sur le site de Sorbonne Université :</p> <p>LERMA, Sorbonne Université, Observatoire de Paris, Université PSL, Cergy Paris Université, CNRS, 75014 Paris [ou 92190 Meudon], France</p> <p>Pour les équipes hébergées sur le site de Cergy Paris Université :</p> <p>LERMA, Cergy Paris Université, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 75014 Paris [ou 92190 Meudon], France</p>
Laboratoire d'études spatiales et d'instrumentation en astrophysique	LESIA, Observatoire de Paris, Université PSL, Sorbonne Université, Université de Paris, CNRS, 92190 Meudon, France
Laboratoire Univers et Théories	LUTH, Observatoire de Paris, Université PSL, Université de Paris, CNRS, 92190 Meudon, France
Station de Radioastronomie de Nançay	USN, Observatoire de Paris, Université PSL, Univ Orléans, CNRS, 18330 Nançay, France
Systèmes de Référence Temps-Espace	<p>SYRTE, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 75014 Paris, France</p> <p>Pour les recherches qui relèvent du périmètre LNE :</p> <p>LNE-SYRTE, Observatoire de Paris, Université PSL, Sorbonne Université, CNRS, 75014 Paris, France</p>

Research data object	Type	Format	Estimated volume	Coverage	Description	Production	Versioning	Licence	Embargo period	Naming rules	Observatory / Space mission / Laboratory	Instrument name
Cassini/RPWS/HFR level 2 collection	data files	binary stream	400 GB	1997-08 to 2017-09	Cassini RPWS/HFR level 2 (measurement in physical values) data collection in hourly files.	basic calibration to convert measured ADUs (level 1) into V^2/Hz	V1.0	CC-BY-NC-SA	none	PYYYYDDD.HH with: YYYY = year DDD = day of year HH = hour of day	Cassini	RPWS
Cassini/RPWS/HFR level 3 Thermal Noise analysis collection	data files	CDF files	60 MB	2004-07 to 2017-09	In-situ plasma density and temperature measured by Cassini/RPWS/HFR during each perikrone, using the Quasi Thermal Noise (QTN) analysis.	Quasi Thermal Noise analysis on Cassini/RPWS/HFR level 2 data	V1.0	CC-BY-NC-SA	none	co_rpws_hfr_qtn_YYYYMMDDHHMM_YYYYMMDDHHMM_vVV.cdf with: YYYYMMDDHHMM (1st instance) = start year-month-day-hour-minute YYYYMMDDHHMM (2nd instance) = end year-month-day-hour-minute VV = version number	Cassini	RPWS
NDA/Routine Jupiter RT1 collection	data files	binary stream	23MB / day	1999-09 to now	Nancay Decameter Array Jupiter Routine observations in raw (RT1) data format	Output from NDA/Routine receiver	V1.0	CC-BY-NC-SA	last month of data is private	JYYMMDD.RT1 with: YY = two digit year number MM = month DD = day	Station de Radioastronomie de Nançay	Nançay Decameter Array
NDA/Routine Jupiter CDF collection	data files	CDF files	23MB / day	1999-09 to now	Nancay Decameter Array Jupiter Routine observations converted in CDF format	Automated conversion with maser4py library	V1.3	CC-BY-NC-SA	last month of data is private	srn_nda_routine_jupiter_edr_YYYYMMDDHHMM_YYYYMMDDHHMM_vVV.cdf with: YYYYMMDDHHMM (1st instance) = start year-month-day-hour-minute YYYYMMDDHHMM (2nd instance) = end year-month-day-hour-minute VV = version number	Station de Radioastronomie de Nançay	Nançay Decameter Array
NDA/Routine Jupiter PDF quicklooks collection	preview files	PDF	300kB / day	1999-09 to now	Nancay Decameter Array Jupiter Routine observations preview in PDF format	NDA/Routine pipeline	V1.0	CC-BY-NC-SA	last month of data is private	JYYMMDD.pdf with: YY = two digit year number MM = month DD = day	Station de Radioastronomie de Nançay	Nançay Decameter Array
NDA EPNcore metadata catalogue	catalogue	PgSQL table	3 new records per day	1999-09 to now	Metadata catalogue for all NDA research data objects produced by the NDA team, following the EPNcore data model.	ingestion into DaCHS server (http://vogate.obs-nancay.fr)	V1.0	CC-BY-NC-SA	none	nda.epn_core	Station de Radioastronomie de Nançay	Nançay Decameter Array