H2020-INFRAIA-2019-1

Europlanet 2024 RI has received funding from the
European Union's Horizon 2020 Research and Innovation Programme under

Grant agreement no: 871149

Deliverable D1.11

Deliverable Title:	2nd Data Management Plan
Due date of deliverable:	31/03/2022
Nature[1]:	ORDP
Dissemination level[2]:	Public
Work package:	WP1,2,3,5,6,7,8,9,10,12
Lead beneficiary:	UniKent
Contributing beneficiaries:	JacobsUni, UniPd, Obsparis, IWF-OEAW, VUA, CNRS, VU
Document status:	Final

Start date of project:	01 February 2020
Project Duration:	48 months
Co-ordinator:	Prof Nigel Mason, UniKent

1. **Nature:** R = Report, P = Prototype, D = Demonstrator, O = Other
2. **Dissemination level:**

PU	PP	RE	CO
Public	Restricted to other programme participants (including the Commission Service)	Restricted to a group specified by the consortium (including the Commission Services)	Confidential, only for members of the consortium (excluding the Commission Services)

Table of Contents

1 Executive summary

The Europlanet H2024 RI Data Management Plan (DMP) is provided. Described aspects in this DMP include: beneficiaries producing data, adherence to FAIR principles, data types, formats and standards, metadata, documentation, intellectual property and data storage, archiving and curation during and after the project.

2 Acronym Table

ASCIII	American Standard Code for Information Interchange
ASI	Agenzia Spaziale Italiana
CC-BY	Creative Commons Attribution (license)
CRISM	Compact Reconnaissance Imaging Spectrometer for Mars
CRS	Coordinate Reference System
CTX	Context Camera
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
DTM	Digital Terrain Model
EOSC	European Open Science Cloud
EPN 2024 RI	Europlanet 2024 Research Infrastructure
EPN-TAP	Europlanet Table Access Protocol
EPNcore	Europlanet Core Data Model
EPN-TAP	EuroPlanet Table Access Protocol (TAP)
ESA	European Space Agency
ESAC	European Space Astronomy Centre (ESA establishment)
EU	European Union
FELICS	Fast Efficient & Lossless Image Compression System
FTP	File Transfer Protocol
GDAL	Geospatial Data Abstraction Library
GMAP	Geological Mapping
GNU	GNU's Not Unix (recursive acronym)
GOES	Geostationary Operational Environmental Satellites
GPL	GNU General Public License
GSF	Guest Storage Facility (for ESA ESAC)
H2020	Horizon 2020 (Framework Programme)

HiRISE	High Resolution Imaging Science Experiment
HRSC	High Resolution Stereo Camera
HTTP	Hyper Text Transfer Protocol
IPDA	International Planetary Data Alliance
ISIS	Integrated Software for Imagers and Spectrometers
IVOA	International Virtual Observatory Alliance
JRA	Joint Research Activity
LROC	Lunar Reconnaissance Orbiter Camera
M3	Moon Mineralogy Mapper
MDIS	Mercury Dual Imaging System
MESSENGER	MErcury Surface, Space ENvironment, GEochemistry and Rangin
MLA	Mercury Laser Altimeter
ML	Machine Learning
MOC	Mars Orbiter Camera
MRO	Mars Reconnaissance Orbiter
NA	Networking Activity
NAC	Narrow Angle Camera
OGC	Open Geospatial Consortium
OMEGA	Observatoire pur la Minéralogie, l'Eau, les Glaces et l'Activité
PADC	Paris Astronomical Data Center
PDS	NASA Planetary Data System (archive, organisation, standard)
PSA	Planetary Science Archive (ESA planetary archive)
PUG	Product User Guide (for ESA GSF)
RDR	Reduced Data Record
RDA	Research Data Alliance
SPASE	Space Physics Archive Search and Extract
SC	Science Case
SPIDER	Sun Planet Interactions Digital Environment on Request

STEREO	Solar TErrestrial RElations Observatory
SME	Small, Medium Enterprise
SNR	Signal to Noise Ratio
SPICE	Spacecraft ... Planet ... Instrument ... C-matrix (orientation) ... Events
SWIR	Short-Wave InfraRed (spectral range)
TA	Transnational Access
TAP	Table Access Protocol (VO)
THEMIS	Time History of Events and Macroscale Interactions during Substorms
USGS	United States Geological Survey
VA	Virtual Activity
VESPA	Virtual European Solar and Planetary Access
VA	Virtual Access
VESPA	Virtual European Solar and Planetary Access
VICAR	Video Image Communication And Retrieval
VNIR	Visible and Near Infrared (spectral range)
VO	Virtual Observatory
WAC	Wide Angle Camera
W3C	World Wide Web Consortium
WP	Work Package
UI	User Interface
URL	Uniform Resource Locator

3 Introduction

The initial version of the Europlanet H2024 RI Data Management Plan (DMP) is provided for relevant data-producing work-packages of the Infrastructure across VA, TA and NA activities.

Described aspects in this DMP include: beneficiaries producing data, adherence to FAIR principles, data types, formats and standards, metadata, documentation, intellectual property and data storage, archiving and curation during and after the project.

Updates of the present DMP will be provided at relevant milestones. The document itself will be live and it will keep evolving throughout the lifetime of the infrastructure.

4 VA

Data management related aspects of Virtual Access activities and related WP, as well as associated JRA are described below.

4.1 VESPA

Work Package Beneficiaries and Partners	
Obsparis	Observatory of Paris, France
JacobsUni	Jacobs University Bremen,
DLR	Deutsches Zentrum für Luft- und Raumfahrt, Germany
CBK-PAN	Centrum Badan Kosmicznych Polskiej Akademii Nauk Poland
CNRS	Centre Nationale de la Recherce Scientifique, France
SpaceFrog	Space Frog France
INAF	Istituto Nazional di AstroFisica, Italy
IWF/OEAW	Oesterreichische Akademie Der Wissenschaften, Austria
UPV/EHU	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea, Spain
IASB/BIRA	Institut Royal D'aeronomie Spatialede Belgique, Belgium
UCL	University College London, United Kingdom

4.1.1 Introduction

VESPA is a virtual research infrastructure. It is fed by contributors, who are sharing their science datasets using VESPA. We define the *VESPA contributors* as all EPN2024RI beneficiaries from TA, NA and VA work packages, as well as any external team willing to share data using VESPA.
We define several types of data products and collections in the frame of this document:

- The *scientific datasets* are composed of individual *data products* and are the result of a scientific observation or analysis produced by each contributor.
- The *metadata catalogues* are derived by *VESPA contributors* from *scientific datasets* and are shared in a common format (EPNcore).

- The *documentation* covers any type of document (reports, user guides, tutorials…)
- The *software* is developed to build and support the infrastructure and serve the science needs of the community.
- Groups of data products (as listed here) can be gathered as *data collections*.

The contributors are generating *scientific datasets* and associated *metadata catalogues*, forming a VESPA *data service,* as shown on Fig. 1. The *metadata catalogues* are the backbone of VESPA. All the *metadata catalogues* are searchable through VESPA query interfaces (either the main VESPA query portal at http://vespa.obspm.fr, or query interfaces embedded in analysis tools). In order to support and facilitate the access to *metadata catalogues* and *scientific dataset* , VESPA beneficiaries are developing *software* . VESPA is also producing *documentation* supporting contributors and users.

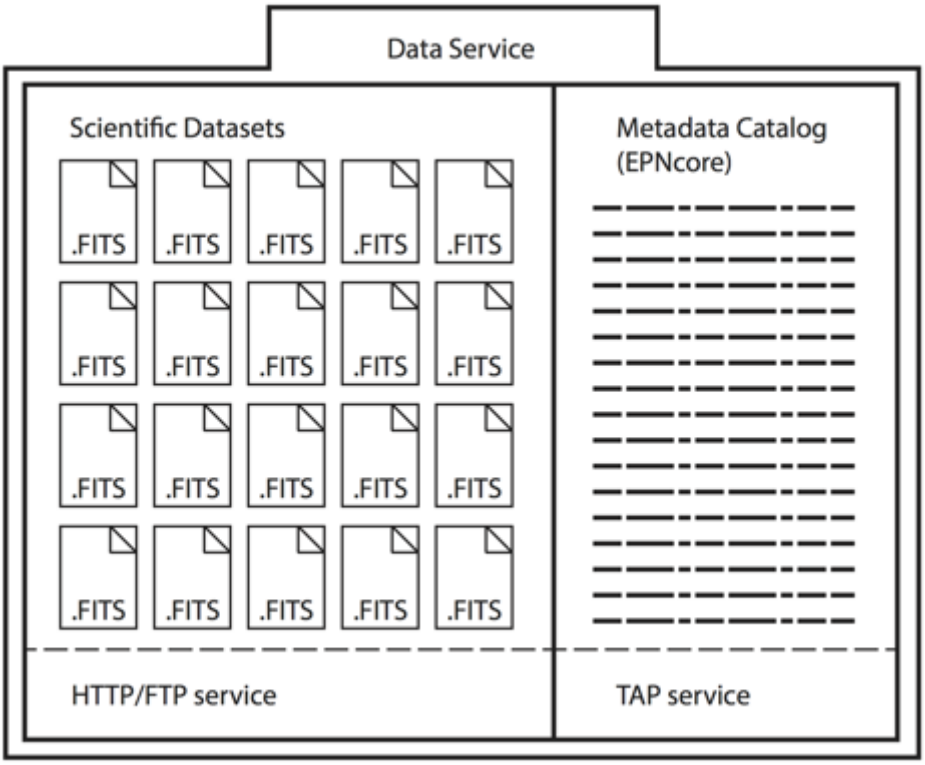

Figure 1. The contributors are generating *scientific data products gathered into a datasets*, using standard file formats (here FITS files) and associated *metadata catalogues* (with the EPNcore specification). The data product included in the *scientific datasets* are shared with open access URLs (using HTTP and/or FTP). The *metadata catalogues* are shared using TAP. This is forming a VESPA *data service.*

4.1.2 Data summary

4.1.2.1 *Scientific datasets*

The *scientific datasets* generated by the EPN2024RI beneficiaries of the VESPA work packages are provided in Annex A. External contributors are not committed by this

document, but it is recommended that they comply with the good practices in terms of data management, as detailed in this document.
The *scientific datasets* are shared using standard and open data formats, as well as standard or commonly used conventions and data models (i.e. metadata formats). The data format defines the overall grammar used to build the file, and the data models are the vocabulary used to describe and write the content of the data product. Both are required for interoperability: the former ensures the data can be read easily, the latter ensures that the data can be interpreted easily. The data products should then include a header containing metadata compliant with a community standard. A list of recommended data formats and metadata standards is available in Annex B. When data have previously been produced in another format, a new dataset may be generated in a recommended format. Both datasets (previous format and recommended format) are published, ensuring continuation of previous access methods and better interoperability.

4.1.2.2 *Metadata catalogues*

The *metadata catalogues* are catalogues of all data products included in each *scientific dataset*. The *metadata catalogues* are built by the contributors from their *scientific datasets.* They contain coverage (temporal, spectral, positional, etc, intervals), provenance and access information for each data product of a *scientific dataset.* All EPN2024RI beneficiaries generating *scientific datasets* are expected to produce a *metadata* catalogue describing each of their *scientific datasets*.
The *metadata catalogues* must be compliant with the EPNcore specification. The EPNcore data model is using IVOA standards. This enables data sharing and discovery within VESPA. The homogeneity of the *metadata catalogues* ensures interoperability. The EPNcore data model has been defined during the Europlanet-RI FP7 and EPN-2020-RI projects. In the course of the EPN2024RI program, EPNcore has been extended. It will be extended in the future when new needs are identified. Extensions are built after assessing pre-existing standards such as IVOA, OGC, RDA, SPASE, and of course W3C, in order to comply with international standards.
All *metadata catalogues* must be checked against a service validator developed by OBSPARIS. Each *metadata catalogue* has an IVOA resource identifier. *Metadata Catalogues* must be compliant with the EPNcore version 2 specification. It is shared using the Table Access Protocol (TAP). Each *VESPA contributor can host and maintain his VESPA distribution server, or rely on the VESPA Hubs.*

4.1.2.3 *Documentation*

VESPA *documentation* consists of various types of documents:

- Assessment studies: working documents discussing the future developments or standard adoptions.
- Reports : summary of activities.
- Recommendations: good practices to *VESPA contributors*, and are submitted for endorsement/adoption as standards to international consortia (IAU, IVOA, IPDA).
- Tutorials: step by step instructions to carry out tasks for *VESPA contributors* and *VESPA users*.
- Other types of documents can be added in the course of the project.

Documentation is internally reviewed and submitted to the VESPA VA review board panelists. Recommendations are reviewed by international consortia (IAU, IVOA, IPDA). Tutorials may be modified and corrected before and after each tutorial session, according to users feedback. VESPA *documentation* follows EC and EPN2024RI quality guidelines.
Documentation is prepared using a document template in the VESPA wiki. Released documents are available in PDF on the main VESPA website.
Documents are released in Plain text, PDF, web content (e.g., HTML or MarkDown) or standard video formats (e.g., MPEG).

4.1.2.4 *Software*

VESPA *software* are tools and libraries. They are developed to enhance the access to *scientific datasets* and *metadata catalogues*. VESPA-JRA are in charge of developing tools that improve the science return of datasets of their discipline, e. g., by enabling VESPA query capabilities in thematically specialized tools.
VESPA *software* is developed using a GIT repository, enabling versioning and backup of codes. Good coding practices (such as automated testing procedures, commenting) are recommended. Open licenses improve maintainability. Modularity and frequent comments improve code understandability for future developers.
VESPA *software* is developed in many languages (Python, Java, Javascript, etc).

4.1.3 FAIR data - findable, accessible, interoperable, re-usable

The virtual observatory concepts and tools are rather new to the planetary science community. Practical training is thus needed for this community to use it and change its habits. For instance, the heliophysics and planetary magnetosphere scientific communities have long been using interoperable tools for science. We have to extend such work methods to the general planetary science fields. The VESPA infrastructure is the way to enter a new interoperability era. The VESPA added value arises from the ability to discover many data products from several data repositories, thanks to the use of a common data model (EPNcore). The VESPA infrastructure is currently openly available from the main VESPA query portal (http://vespa.obspm.fr). Dedicated accesses are developed in specific open tools, such as spectroscopy analysis tools, or planetary surfaces plotting interfaces, as well as in processing software such Python.
All "open access" and licensing statements in the following subsections hold unless other binding agreements exist.

4.1.3.1 *Scientific datasets*

Scientific datasets are open access.
Access may be reserved during a proprietary period, so that teams can work on the data and prepare their service in advance, before releasing the data. This allows the use of VO access for collaborative work even during the preparation of a project.
All scientific datasets are distributed with CC-BY license (unless otherwise specified), with a possible reference to a document indicating the rules of use (citation and acknowledgement policy).

4.1.3.2 *Metadata catalogues*

All *metadata catalogues* are open access. The VESPA project is focusing on enhancing the visibility and return on the *scientific datasets*. The main audience for VESPA is the scientific community. However, the access to *metadata catalogues* through VESPA can reach a wider audience, such as education, public outreach, amateur astronomers, etc. Online tools enabling an easy query, access and visualization of *scientific datasets* are developed, or reused and improved when already existing. Basic tools proposed to VESPA users are known for the reliability and experience of the development teams, which are mostly outside the EPN2024RI project. All the tools are developed with open source licenses, ensuring security, stability and sustainability.

The *metadata catalogues* include a scientific description of the data. Any scientific team may thus immediately select data products from *scientific datasets*. Any software client can also connect to, ingest and compare data from several *metadata catalogues* and *scientific datasets* at the same time. The EPNcore *metadata catalogues* are based on specific standards, ensuring the sustainability of the infrastructure. Every data product described in a VESPA *metadata catalogues* is reachable through an open access URL. Private data distribution (requiring authenticated access) are possible during embargo periods for recent data.

VESPA is using the IVOA Registry to collect and maintain the list of available services providing *metadata catalogues*. The IVOA Registry service is maintained by IVOA and PADC (OBSPARIS), independently from the VESPA project. The Registry is replicated in Europe and in the USA, so that there is no 'single point failure'. All VESPA *metadata catalogues* services are independent, preventing global failures.

4.1.3.3 *Documentation*

A Document Register has been set up to manage documentation internally. Released versions of documents are available from the VESPA main website. A copy of this document register at time of writing is present in Annex A.

VESPA *documentation* is licensed under CC-BY (unless otherwise specified).

4.1.3.4 *Software*

All software developed in the frame of VESPA activity are open source. The GPLv3 license is strongly recommended for codes developed in the project. In case of pre-existing software bricks, already protected under another license, licensing should be defined on a case by case basis, selecting a license endorsed by the Open Source Initiative. The VESPA coordination encourages and helps teams to identify a software license for their existing codes if there was none.

VESPA software are openly accessible from the VESPA GIT online repository. Other open repositories existing in VESPA beneficiary institutes are also used for code sharing.

4.1.3.5 *Persistent Identifiers*

VESPA products and collections defined previously are made accessible and findable thanks to a persistent identifier. A DOI (Digital Object Identifier) shall be associated to any reusable and citable data product. A document describing the DOI good practices for VESPA is available (Cecconi et al, 2020, doi:10.25935/p90x-ty59). See this document for more details.

4.1.3.6 *Individual Data Management Plan*

Each collection will be managed independently with a dedicated DMP document, which will follow the template provided by VESPA available from this page: https://voparis-wiki.obspm.fr/display/VES/Individual+DMP+of+EPN2024RI+collections.

4.1.3.7 *Short term storage and data management*

VESPA is a distributed system where contributors host and maintain their own data services. Metadata catalogues can point to data in existing archives, e.g. at the ESA/PSA, therefore providing easier access to a subset of data of interest. Each *VESPA contributor* is in charge of hosting and storing their own datasets (*metadata catalogues* and possibly *scientific datasets*).

VESPA is enabling the sharing of *metadata catalogues* to provide a better access to *scientific datasets*. The datasets are distributed. There is no plan to have a centralized repository for datasets, as it would be contradictory with the concepts of the Virtual Observatory, where each contributor is in charge of providing the latest version of its datasets. The VESPA project is not responsible for the sustainability of the individual datasets, but for that of the infrastructure. Each *VESPA contributor* is then responsible for the sustainability of its own databases. However, *VESPA Hubs* may take over hosting and distribution tasks, upon request of the *VESPA contributor*, and provided technical feasibility; usage of the VESPA Hubs is encouraged to all data providers.

The VESPA infrastructure sustainability is ensured in the short term through a VESPA helpdesk, and on the long term by having VESPA developments and standards endorsed by large scale international organizations (IAU, IVOA and IPDA). A data management support is also provided to *VESPA contributors* through the VESPA helpdesk (support.vespa@obspm.fr).

4.1.3.8 *Curation and preservation*

In a Virtual Observatory paradigm the data are hosted and delivered by the scientific teams, who have the scientific capabilities and knowledge to ensure accurate description of the data products and update them when necessary. Each *VESPA contributor* then acts as a Virtual Observatory node feeding the infrastructure with datasets. The curation and preservation policies rely upon each *VESPA contributor*.

The curation of VESPA datasets is conducted by contributors on their own datasets. The VESPA datasets are reviewed by the VESPA team, so that their content meets the quality level allowing efficient science queries (e.g, checking thumbnail quality, keyword consistency, custom keywords). VESPA provides a validator for assessing the conformance of the service with respect to VESPA standards. All available services can also be tested at once for monitoring purposes. This includes the updates of datasets and the selection of documents and versions of datasets to be kept for preservation. Annex C provides the preservation plans for each VESPA beneficiaries.

A specific care is taken with datasets generated by the amateur community. Such data products are validated before being shared with the scientific community through VESPA. The VESPA team conducts the data validation. Rejected amateur

data products are not stored by the hosting institutes. Curation and preservation of the validated data follow the same policies as other datasets and rely on each VESPA contributor dealing with amateur datasets.
In case *VESPA contributors* cannot commit to long term preservation of their datasets, usage of community-based preservation platforms such as Zenodo (or alike, e.g. Europlanet community on B2SHARE) is recommended.

4.2 GMAP

Work Package Beneficiaries and Partners	
Unipd	University of Padova, Italy
JacobsUni	Jacobs University Bremen, Germany
DLR	Deutsches Zentrum für Luft- und Raumfahrt, Germany
INAF	National Institute for Astrophysics, Italy
CBK PAN	Centrum Badan Kosmicznych Polskiej Akademii Nauk
UDA	Universitá d'Annunzio, Italy

4.3 Introduction

The present section describes the type of data, their characteristics and their use, archiving and preservation plans throughout the GMAP project. The section outlines the basic data management directions that are going to be updated throughout the project and issued at discrete steps.

4.4 Beneficiaries using data

All beneficiaries will use data, in either individual or - in most cases - combined forms. Data access to archived mission (NASA/ESA) data is free for anyone. Some data will have temporary team-only access (during an embargo of up to several months), such as mapping data used by GMAP researchers (see section on Data Storage and Management during the project)

4.4.1 Beneficiaries producing data

All beneficiaries will produce either derived data (higher-level data products) or new data derived by both human and computer/algorithm-assisted mapping.

4.4.2 FAIR data - findable, accessible, interoperable, re-usable

Data produced by GMAP will impact future robotic and space exploration, mainly through mature, finished, published mapping products. Underlying data and special mapping products will be of scientific use also before and beyond that. Accessibility to the data will be provided in three different forms:
Findable data:

- Longer-term discoverability will be guaranteed via connected Institutional repositories (ESA, UNIPD, INAF), VESPA sharing and inclusion in well-established planetary data archives that are accessible and commonly used by the community.
- Shorter-term discoverability will be supported by the GMAP data-hub access

Accessible data:

- Geological mapping products will be finalized in self-contained and reproducible packages, downloadable over the internet from the GMAP data-hub website or other storage services. The dataset will be made freely available to the public after package finalization. During the GMAP project availability might be restricted to individual units for dataset intended for internal use or not yet finalized.
- Interoperable data:
  - OGC standards for CRS and formats will be adopted
  - Data discovery interoperability will be granted via the use of state-of-the-art VESPA EPN-TAP (Virtual European Solar and Planetary Access EuroPlanet Table Access Protocol) for data search and query.
- Re-usable data:
  - Custom base-map data (e.g. mosaics) and partial mapping products and processed/derived datasets underlying geological mapping products (standard, non-standard, integrated, etc.) will be usable by others, also in the future, regardless of the final geological mapping products.

Integrated and/or final mapping products will be re-usable directly or indirectly, with access to combined information content or individual layers with relevant topologies, (units, contacts, etc.).

4.4.3 Data types, formats and standards

GMAP uses existing datasets and data products and creates new products deriving from combination or derivation of existing, processed data products, as well as from completely new mapping (e.g. geological or geomorphological units),

4.4.4 Data

4.4.4.1 *Raw data*

Planetary archives, PDS3, PDS4 imagery and cubes.

4.4.4.2 *Base mapping data*

OGC-compliant data already available from external entities (e.g. USGS) or base mapping data produced by GMAP partners, some in PDS standards/formats.

4.4.4.3 *Integrated mapping products*

Integrated mapping products with individual layers are being produced in OGC-compliant formats, both raster and vector, as well as with suitable 3D formats (See Annex D). All individual layers/components of maps are in geospatial-aware formats and with CRS suitable for the specific mapping project: in-situ, local, regional or global derived mapping products will be possibly generated.

4.4.5 Metadata

The aim of including metadata is to allow reproducibility and reusability by providing information about the processing steps performed and the properties of the mapping product which are fundamental for adequate reuse of the map. Map-wide metadata including both geometric and bibliographic information are going to be provided for each map.

.

4.4.5.1 *Raw data*

Metadata from processed raw data are the same as those from archived data. SPICE kernel version and software used (e.g. USGS ISIS) should be recorded.
Isis Cube labels (i.e. recording cumulative processing steps and used ancillary data, metadata, CRS and alike). The information is going to be recorded in processing labels and/or as temporary output in ASCII format.

4.4.5.2 *Base mapping data*

Projections, cubes and images used, type of control network used, and relevant additional information available from original derived data producers (e.g. USGS, ESA, academic institutions or local GMAP base mapping data producers or groups) will be recorded.

4.4.5.3 *Integrated mapping products*

Metadata for integrated mapping products will be both map-related and sub-map-related (i.e. per geological unit description).
Map-related metadata include, as a minimum:

- Used datasets and products
- Mapping individuals
- CRS
- Summary of used tools and documented workflow

Unit-related information will be recorded and updated during the mapping process and will include, as a minimum:

- Individual products and layers used to determine unit contacts
- Eventual interpolation/extrapolation of data underlying mapped unit outline
- Qualitative assessment on uncertainties involved in the unit determination

Information about the authors of the maps, programs used and processing steps will be included and added to the documentation to allow reproducibility of the underlying workflow .

4.4.6 Map types

GMAP maps will be named by using a coded designation letter, for each map type:

- **S** = Stratigraphic
- **C** = Compositional
- **M** = Morphologic
- **G** = Geo-structural
- **I** = Integrated
- **D** = Additional DOM-specific mapping products for individual or multiple lander/rover-imaged outcrops can be included

4.4.6.1 *Map-level metadata*

Complementing metadata related to individual units, each map of GMAP include several map-wide fields, exemplified below:

Field	**Field description (and example entries)**
Map name (PM_ID)	GMAP-MER-MS-H02_3cc
Target body	Mercury
Title of map	Geologic Map of the Victoria Quadrangle (H02), Mercury
Bounding box - Min Lat	-22.5°
Bounding box - Max Lat	65°
Bounding box - Min Lon (0-360)	270°
Bounding box - Max Lon (0-360)	360°
Author(s)	Valentina Galluzzi; Laura Guzzetta; Luigi Ferranti; Gaetano di Achille; David A. Rothery; Pasquale Palumbo
Type	Released
Output scale	1:3M

Original Coordinate Reference System	Lambert conformal conic Center longitude: 315° Standard parallel 1: 30° Standard parallel 2: 58° Datum: 2440 km (non-IAU, MESSENGER team datum)
Data used	MESSENGER MDIS BDR v0 uncontrolled basemap (166 m/pixel) MESSENGER MDIS 2013 complete uncontrolled basemap (250 m/pixel) MESSENGER MDIS uncontrolled mosaics v6, v7, v8 (250 m/pixel) MESSENGER MDIS partial mosaic (USGS) (200 mpp) MESSENGER MDIS 2011 albedo partial mosaic (USGS) (200 m/pixel) Mariner 10 + MESSENGER flyby uncontrolled basemap (USGS) (500 m/pixel) MESSENGER MLA DTM (665 m) MESSENGER MDIS M2 flyby stereo-DTM (DLR) (1000 m)
Standards adhered to	Mapping scale: Tobler (1987); Output scale: USGS; Symbology: USGS FGDC and other new symbols
DOI	10.1080/17445647.2016.1193777
Aims	Morpho-stratigraphic analysis of Mercury's units and BepiColombo target selection.

Short description	Mercury's quadrangle H02 'Victoria' is located in the planet's northern hemisphere and lies between latitudes 22.5° N and 65° N, and between longitudes 270° E and 360° E. This quadrangle covers 6.5% of the planet's surface with a total area of almost 5 million km2. Our 1:3,000,000-scale geologic map of the quadrangle was produced by photo-interpretation of remotely sensed orbital images captured by the MESSENGER spacecraft. Geologic contacts were drawn between 1:300,000 and 1:600,000 mapping scale and constitute the boundaries of intercrater, intermediate and smooth plains units; in addition, three morpho-stratigraphic classes of craters larger than 20 km were mapped. The geologic map reveals that this area is dominated by Intercrater Plains encompassing some almost-coeval, probably younger, Intermediate Plains patches and interrupted to the north-west, north-east and east by the Calorian Northern Smooth Plains. This map represents the first complete geologic survey of the Victoria quadrangle at this scale, and an improvement of the existing 1:5,000,000 Mariner 10-based map, which covers only 36% of the quadrangle.
Related products	Geologic Map of the Hokusai Quadrangle (H05), Mercury Geologic Map of the Shakespeare Quadrangle (H03), Mercury Geologic Map of the Kuiper Quadrangle (H06), Mercury

Units Definition (polygon styling)	Smooth Plains, sp, 255-190-190 Northern Smooth Plains, spn, 245-162-122 Intermediate Plains, imp, 245-122-122 Intercrater Plains, icp, 137-90-68 Crater material-well preserved, c3, 255-255-115 Crater material-degraded, c2, 92-137-68 Crater material-heavily degraded, c1, 115-0-0 Crater floor material-smooth, cfs, 255-255-175 Crater floor material-hummocky, cfh, 205-170-102
Stratigraphic info	This map has an associated database of craters larger than 5 km used for basic crater frequency analysis for N(5), N(10), and N(20).
Other comments	Since the mapping scale (~1:400k) was much higher than the output scale (1:3M) the polylines of the map were not smoothed. This map is currently being updated to fit the new controlled MESSENGER's end-of-mission basemaps. A post-release boundary merging was done with the H03 and H05 quadrangles. This map uses a legend also for feature labels.
Heritage used	former Mariner 10 map by McGill and King (1983)
Link to other repositories	(crater database link) (shapefiles database link)
Acknowledgements beyond GMAP	This research was supported by the Italian Space Agency (ASI) within the SIMBIOSYS project [ASI-INAF agreement number I/022/10/0]. Rothery was funded by the UK Space Agency (UKSA) and STFC.

Table 1: Exemplary map-wide metadata for GMAP products, exemplified by an existing PLANMAP in-kind contribution (Geologic map by Galluzz i et al., 2016).A subset of metadata are (see section on ESA PSA delivery) is also included on the datasets released to ESA.

4.4.7 Documentation

Documentation of GMAP will be available on the project wiki space (wiki.europlanet-gmap.eu, which will be kept functional after project-end based on best effort and availability of resources. The internal wiki space is used for both internal project coordination and technical, scientific documentation. The latter, in evolved form, will be also shared via the project public wiki space.
The types of documentation in the GMAP wiki include:

- Summary of relevant activities per WP
- Procedures and workflows
- Mapping use case description
- Best practices and recommendations
- Tutorials on data handling and mapping
- Other documents

4.4.8 Software

The software used to access and analyze GMAP data will be based on Open Standards, in particular OGC standards. Both Open Source and proprietary software (such as QGIS, ArcGis) will therefore be suitable for accessing the data. A particular case is constituted by the software that is employed for 3d geological modeling, for which open source alternatives rarely exist. For the choice of the software package two criteria will be considered: a) the feasibility for the task that will be undertaken b) the academic licensing scheme that is adopted. Under the same feasibility conditions, software packages granting low-cost/affordable licensing schemes for academic purposes will be favoured.
The consortium will use a wide range of publicly available Open Source and commercial tools to work and perform mapping tasks. Additionally, algorithmic and programmatic methods that add value to interactive human-computer mapping will also use, as far as possible, Open Source tools, packages and libraries.
Software, tools and scripts or snippets developed throughout the project will be shared both internally and externally via the organization‘s GMAP GitLab page (gitlab.com/geo-bl-ch/gMap). Some repositories might be private, with access restricted to beneficiaries, during the early phases of the project. Ultimately, all will be made public and will be made available indefinitely after the end of GMAP.

4.4.9 Data exploitation, accessibility and intellectual property

Intellectual property rights on individual science outputs will be held by the scientific collaborators and publishing venue/journal (e.g. individual papers). Data and maps published on the GMAP data archive (ESA Guest Storage facility, INAF, UNIPD or other institutional data repositories) and their long-term evolution are cited either via their dataset DOI or via relevant linked publications.

4.4.9.1 *Data and metadata*

Produced base mapping data are provided as CC-BY (attribution).

Published maps (of any kind) are going to be provided, free to use, with CC-BY (attribution).
Acknowledgment of the GMAP Europlanet H2024 RI is requested from those using GMAP-derived data. A relevant acknowledgement message will be included in the documentation provided to ESA, as well as within the global metadata of VESPA-shared datasets.

4.4.9.2 *Documentation*

Documentation licensing will follow Creative Commons CC-BY-4.0 (https://creativecommons.org/licenses/by/4.0/). Documentation will be also available, complementing or copying information on the public wiki space, on GitLab and possibly other public repositories.
At the end of the project the entire body of documentation will be consolidated and available both on the GMAP public wiki (wiki.europlanet-gmap.eu) and GitHub/GitLab (gitlab.com/geo-bl-ch/gMap).

4.4.10 Software

Software developed by GMAP partners is going to be open source. License GPLv3 is recommended, or any other license covered by the Open Source initiative (https://opensource.org/licenses/category).
Software, tools and scripts produced by GMAP will be available as soon as they are considered usable, on the public GitLabor GitHub page (github.com/europlanet-gmap). Private repositories will be used during the course of the project but will cease to exist at its end and all will be made public.

4.4.10.1 *Data/Software citation*

Archived data used that comes from mission archives will follow the custom of quoting experiment description papers and eventual relevant follow-up papers (e.g. Malin et al., 2007; McEwen et al., 2007; Neukum et al., 2004; Jaumann et al., 2007). Datasets from NASA/ESA archives (PDS, PSA) follow the citation requirements of those archives. In the case of NASA public domain data, the experiment-description papers (e.g. Malin et al., 2007) would be cited in scientific publications. ESA data follow similar citation styles (suggested citations are included in the PSA entry pages).
Datasets produced by the GMAP consortium will be possibly quoted via:

- Relevant peer-reviewed publications or published maps indicated in the dataset metadata (similar to PSA/VESPA)
- Dataset-specific DOI, i.e. via OpenAIRE/Zenodo/GitHub for relevant datasets
- Eventual additional DOI-generating data services that might become available during the project lifetime

4.4.11 Data storage and management during project

During day-to-day operations and technical/scientific activities of the consortium, data will be stored on each partner's premises as well as, when relevant, on shared network resources (such as cloud, FTP and web mapping data access services). In

principle, data will be made publicly available as soon as possible during the project, respecting publication embargoes.

4.4.12 Data curation, archiving, preservation and security

4.4.12.1 *Data curation*

Base data and maps (See Annex A) undergo archiving review by the archive maintainers (PDS, PSA). If any issue is encountered (e.g. missing or problematic labels, metadata and/or eventual problems with data themselves) the GMAP consortium will share that information with respective archive data publishers (PDS, PSA).
Mapping is an iterative, interactive process that will go through a few levels of interactive, informal and formal scientific review within the GMAP partners and consortium. Before a final map is produced (and its related scientific publication is submitted), preliminary versions will be shared on the GMAP web page, wiki and other data-serving web resources. In case underlying base mapping data require or are subject to improvements that will affect the mapping, newer versions will be used and posted, and metadata updated.

4.4.12.2 *Data preservation*

Input data (based data and maps) are preserved by the respective archives and not under GMAP responsibility.
Custom higher-level data imagery, cubes, virtual environments and 3D models produced by GMAP partners during the course of the project will be preserved on GMAP storage services. After the project data will be shared (See Data Sharing subsection), optimally with some redundancy and in different geographic locations, for longer-term availability.

4.4.12.3 *Data security*

The GMAP data processed, produced and analysed are not sensitive. No specific security measures are planned. Data recovery, in case of storage failures will be optimised by the use of central backup and local copies across GMAP partner institutions.

4.4.12.4 *Data sharing*

Data sharing will be performed via 4 possible channels:

- Individual partners e.g. on own web site or repositories using industry standards for geodata (e.g. web-GIS)
- GMAP consortium, via web-gis and data-access web page, linked form the GMAP web page
- ESA PSA, upon delivery of data and mapping products
- VESPA, via distributed VO-compliant systems for integrated mapping products and, in the future, potentially sub-map, mapping unit-level access (e.g. individual mapping units).

4.4.12.4.1 EPN-TAP VESPA-based sharing on premise

VESPA-shared data contain data-product-level metadata pointing to actual data sources. The release of data (see DoA) is planned in steps, to conclude by the end of the project.

4.4.13 Exemplar metadata for mapping products to be released via VESPA:

A set of mandatory, documented metadata for VESPA services exists (https://voparis-confluence.obspm.fr/display/VES/Implementing+a+VESPA+service), plus optional ones. Those are mostly related to data-products (in PDS sense) with some dataset-wide. Individual unit granularity is not yet covered by VESPA technical capabilities.
New developments of VESPA (currently not implemented but envisaged for future VESPA developments within the lifetime of GMAP) should allow for metadata-based discovery and search that could extend the geographic data search and experiment metadata search with feature/unit data/metadata search.

4.4.14 ESA PSA data deliveries

Individual used data products already existing in planetary archives (PDS, PSA) will not be released to PSA (already in PDS in either raw or processed form).

4.4.15 Exemplary metadata for mapping products to be released via PSA:

Release to PSA of non-PDS data, geological mapping data, with relevant documentation is going to be provided. Periodically, shortly after each formal delivery, possibly more often if needed, data from the GMAP webpage are being provided to ESA with additional map-wide metadata and description as required by ESA (ESA, 2019) including:

- target body
- geographic extent (bounding box) of mapping product
- CRS
- additional fields as described in the Map-wide metadata table in the above sections.

Data exchange formats for archived data will include:

- For raster data = preferentially Geotiff
- For vector data = preferentially OGC/Geopackage
- Additional files or the same version of release raster files also provided in different formats might include e.g. ISIS3 cube (.cub) format, Envi (.img + .hdr) or alike.

A copy of the data in either geotiff (raster) and Geopackage (vector) will be provided in any case, where relevant.
Each GMAP map (with a unique GMAP_ID) is considered as a dataset for ESA GSF, with specific authors and DOI. Each map includes typically more products (e.g. pdf layout view, one or more raster imagery, one or more vector datasets packed inside a Geopackage).

The structure of the PUG for each GMAP map / GMAP ESA GSF dataset is following such structure:

1. Introduction
1.1 Dataset (PM_ID) introduction
1.2 Abbreviations and acronyms
1.3 References and applicable documents
1. Dataset generation
2.1 Data source
2.2 Processing steps
2.3 Data product
1. Archive format and content
3.1 Naming convention
3.2 Data types and formats
3.3 Directory structure
3.4 Dataset content
1. Software
2. Caveats and issues

Such PUG is going to be based on GMAP public deliverables and relevant literature therein.

4.5 ML

Work Package Beneficiaries	
ACRI-ST	ACRI-ST, France
AOP	Armagh Observatory and Planetarium, Ireland
DLR	Deutsches Zentrum für Luft- und Raumfahrt, Germany
KNOW	Know-Center GmbH, Austria
IAP-CAS	Institute of Atmospheric Physics, Academy of Sciences of Czech Republic, Czech Republic
INAF	National Institute for Astrophysics, Italy
IWF-OEAW	Space Research Institute, Austrian Academy of Sciences, Austria
LMSU M.V.	Lomonosov Moscow State University, Russia
UNIPASSAU	University of Passau, Germany

4.5.1 Data Summary

4.5.1.1 Introduction

The main objectives of the "JRA4 ML - Machine Learning Solutions for Data Analysis and Exploitation in Planetary Science" work package (ML WP) within the Europlanet 2024 Research Infrastructure (EPN-2024-RI), as stated in the project proposal, are:

- To develop machine learning (ML) tools, designed for and tested on planetary science cases submitted by the community, and to provide sustainable, open access to the resulting products, together with support documentation
- To foster wider use of ML technologies in data driven space research, demonstrate ML capabilities and generate a wider discussion on further possible applications of ML
- To identify scientific and commercial applications for the ML tools developed through the JRA tasks

The goal is to build a multipurpose toolset for ML-based data analysis that will be applicable to a range of scientific research questions in planetary science with minor/ easily-achievable customization efforts. In the course of proposal preparation, the scientific community proposed about a dozen of science cases, of which a subset was selected to be investigated in the ML WP. For the study and analysis of the selected science cases, spacecraft and laboratory data will be utilized. All tools developed and corresponding data products will be linked via the VA services of VESPA, GMAP and SPIDER (where appropriate). An onboarding of the data products and software generated in the ML WP to the European Open Science Cloud (EOSC) is contemplated. Further, the tools will be shared and made accessible to the wider planetary science community through a so-called Machine Learning Portal (ML Portal, https://ml-portal.oeaw.ac.at), which is basically a webspace hosted at the WP lead IWF-OEAW, and will additionally be maintained on a git service (GitHub and GitLab). Whereas the GitLab repositories are private and only for internal developing purposes, the GitHub repositories (https://github.com/epn-ml) are public and for dissemination of all data meant to be offered to the scientific community. The ML WP will also produce documentation and tutorials, which will also be available through the ML Portal and the git service, to support the beneficiaries and the users of the ML tools.

4.5.1.2 Data types and formats

The term d*ata* in this document is defined in a broader sense than commonly used. Data collected as well as generated in this WP are essentially of five different types, specified in the following.

4.5.1.2.1 Original data

For the science cases treated in this WP, data from different sources are used. On the one hand spacecraft data from e.g., Cluster, GOES, MESSENGER, MRO, STEREO, THEMIS, and Van Allen Probes are collected, and on the other hand data from laboratory experiments are generated. The spacecraft data are freely available through the missions' homepages. The laboratory data are provided by the beneficiaries

conducting the experiments. The formats of these data are standard formats, e.g., pds, fits, or cdf.

4.5.1.2.2 Pre-processed data

Data pre-processing and preparation is necessary to transform the original data sets into data sets appropriate for further analyses. The formats of these pre-processed data will be standard formats used throughout the science community, e.g., tiff, png, jpg, csv, or ascii.

4.5.1.2.3 Software

Software developed in the ML WP mean essentially the ML-based tools used to analyze the data of the science cases. The open-source programming language Python will be used for developing these tools. Tools in this regard essentially mean Python scripts, code, and trained models, which are developed to analyze a specific data set for a science case and which can be used and modified by the user to fit own needs. The scripts will include code for data pre-processing, for applying ML algorithms to train models, and for data post-processing. We also refer to this as an ML pipeline.

4.5.1.2.4 Data results

Data results are obtained via applying the ML pipeline to the original or pre-processed data. These data will have standard formats used throughout the science community, e.g., tiff, png, jpg, csv, geojson, shapefile, or ascii.

4.5.1.2.5 Documentation and tutorials

Documentation of the data sets and ML pipelines will be provided. Additionally, tutorials, either in the form of text documents, Jupyter notebooks, videos, or virtual workshops, will be made available. These documents are released in standard formats, e.g., plain text, pdf, or mpeg. The virtual workshops will be conducted through an appropriate platform, e.g., Go2Webinar, or Zoom.

4.5.2 FAIR data - findable, accessible, interoperable, re-usable

A website will be provided by IWF-OEAW, named ML Portal and accessible under *https://ml-portal.oeaw.ac.at*. This website serves as the virtual access point for our activity. The content on the ML Portal will be completely open to anyone interested and it will provide

- an introduction to the ML activity within Europlanet,
- general information about the science caes and their ML challenges,
- links to the Python scripts and codes on a git service,
- documentation and tutorials on how to use the ML pipelines,
- presentations of the results of the science cases,
- announcements of ML related events, e.g., workshops, or conferences.

The scientific data sets, their documentation and the corresponding software will be maintained on a repository of a git service (GitHub and Gitlab). All of these will contain metadata as well as version numbers. Search keywords and Digital Object Identifiers (DOIs) will assist finding the data. Whenever a publication is submitted with a specific version of an ML pipeline, this version will be published with a DOI on Zenodo or a similar service.

After the exploitation of the data products for the science cases and the publication of the results, they will be made freely available. Further, the data products, including the original data sets as well as trained models (where applicable), are aimed to be published in VESPA, using EPN-TAP to describe the data and a TAP service (e.g. DaCHS) to allow VESPA (or other services) to access and harvest the data. The metadata of our data will comply with the VESPA metadata standard.

The data sets, documentation, and tutorials will be published under the CC-BY licenses. Software developed in the course of the ML WP will be open source (Apache License 2.02), thoroughly documented and available via GitHub (https://github.com/epn-ml). These licenses guarantee that all data can be used freely, and can be further developed and extended by the community.

We aim to use standard data formats for all of our data produced in this WP. Thus, all of our data will be interoperable and re-usable by the scientific community.

The ML Portal and the repositories on the git service will be sustained after the end of the project.

4.5.3 Ethical aspects

There are no ethical aspects that need to be considered.

4.6 SPIDER

Work Package Beneficiaries and Partners	
CNRS	Centre National de la Recherche Scientifique, France
OBSPARIS	Observatoire de Paris, France
ONERA	Office National d'Etudes et de Recherches Aérospatiales, France
IRF	Institutet för rymdfysik, Sweden
UCL	University College London, UK
Wigner	Wigner Research Centre for Physics, Hungary
INAF	National Institute for Astrophysics, Italy

4.6.1 Data Summary

The H2020 Europlanet-2020-RI programme, which ended on Aug 31st, 2019, included an activity called PSWS (Planetary Space Weather Services), which provided 12 services distributed over four different domains (A. Prediction, B. Detection, C.

Modelling, D. Alerts) and accessed through the PSWS portal (http://planetaryspaceweather-europlanet.irap.omp.eu/):

A1. 1D MHD Solar Wind Prediction Tool – HELIOPROPA,
A2. Propagation Tool,
A3. Meteor showers,
A4. Cometary tail crossings – TAILCATCHER,
B1. Lunar impacts – ALFIE,
B2. Giant planet fireballs – DeTeCt3.1,
B3. Cometary tails – WINDSOCKS,
C1. Earth, Mars, Venus, Jupiter coupling- TRANSPLANET,
C2. Mars radiation environment – RADMAREE,
C3. Giant planet magnetodiscs – MAGNETODISC,
C4. Jupiter's thermosphere, D. Alerts.

In the framework of the starting Europlanet 2024 RI programme, SPIDER will extend PSWS domains (A. Prediction, C. Modelling, E. Databases) services and give the European planetary scientists, space agencies and industries access to 6 unique, publicly available and sophisticated services in order to model planetary environments and solar wind interactions through the deployment of a dedicated run on request infrastructure and associated databases:

C5. A service for runs on request of models of Jupiter's moon exospheres as well as the exosphere of Mercury,
C6. A service to connect the open-source Spacecraft-Plasma Interaction Software (SPIS) software with models of space environments in order to compute the effect of spacecraft potential on scientific instruments onboard space missions. Pre-configured simulations will be made for Bepi-Colombo and JUICE missions,
C7. A service for runs on request of particle tracing models in planetary magnetospheres,
E1. A database of the high-energy particle flux proxy at Mars, Venus and comet 67P using background counts observed in the data obtained by the plasma instruments onboard Mars Express (operational from 2003), Venus Express (2006–2014), and Rosetta (2014–2015),
E2. A simulation database for Mercury and Jupiter's moons magnetospheres and link them with prediction of the solar wind parameters from Europlanet-RI H2020 PSWS services,
A1. An extension of the Europlanet-RI H2020 PSWS Heliopropa service in order to ingest new observations from Solar missions like the ESA Solar Orbiter or NASA Parker Solar Probe missions and use them as input parameters for solar wind prediction.

Runs on request services will be made available in the web with forms to enter the input parameters to be filled online.

Data from SPIDER databases (E1, E2, A1) will have standard formats for spacecraft missions, e.g., pds, fits, or cdf.

Data from run on request services (C5, C6, C7) will have standard formats commonly used throughout the science community, e.g., cdf, fits, tiff, png, jpg, csv, or ascii.

Documentation and tutorials will be released in standard formats, e.g., plain text, pdf, or mpeg.

4.6.2 FAIR data - findable, accessible, interoperable, re-usable

A website will be provided by CNRS, accessible under http://planetaryspaceweather-europlanet.irap.omp.eu/. This website serves as the virtual access point for SPIDER activities. Its content is completely open to anyone interested and it will provide

1. an introduction to the SPIDER activities within Europlanet,
2. links to SPIDER tutorials and workshops,
3. documentation and tutorials on how to use the services.

The scientific data sets, their documentation and the corresponding software will be maintained on a repository of a git service (e.g. github or Gitlab). All of these will contain metadata as well as version numbers. Search keywords and Digital Object Identifiers (DOIs) will assist finding the data products.

SPIDER data will be published in VESPA, using EPN-TAP to describe the data and a TAP service (e.g. DaCHS) to allow VESPA (or other services) to access and harvest the data products. The metadata of our data products will comply with the VESPA metadata standard. An architecture for publishing the services generated by SPIDER in the European Open Science Cloud (EOSC) will be studied and prototyped during the project.

The data sets, documentation, and tutorials will be published under the CC-BY licenses. Software developed in the course of the ML WP will be open source, thoroughly documented and available via a git service. This guarantee that all results can be used freely, and can be further developed and extended by the community.

The SPIDER website and the repositories on the git service will be sustained after the end of the project.

4.6.3 Ethical aspects

There are no ethical aspects that need to be considered.

5 Transnational Access

5.1.1 *Data Summary*

A set of protocols for the open access storage of data collected during TA visits has been agreed and set up in consultation with the host laboratories and field sites. This was accomplished under guidance of the IT scientific support department at VUA working in close cooperation with VESPA. TA visitors will be granted a year's embargo to providing the data after their TA visit during which time they will be encouraged to publish their data in open access journals and make presentations at international conference. The exceptions to the one year embargo are: i) PhD students who have yet to complete their thesis; ii) commercially sensitive data generated by industrial partners. A year after completion of the TA visit, the relevant scientist is required to supply the data obtained during the TA visit to the TA office at the VUA. Signature of an agreement to provide the data is part of the protocol in accepting the funding for a TA visit.

5.1.2 FAIR data - findable, accessible, interoperable, re-usable

TA visitors will be provided with guidelines on how to prepare the metadata and documentation of datasets that will provide clear data description and allow effective data storage and integration of different types of data related to a specific TA visit; e.g., field pictures and subsequent sample analysis. These guidelines have been developed to ensure that data storage will follow the FAIR principle while maintaining user friendly capability for the host TA labs and users; i.e. deliver of data in standard tabular and text formats that are compatible with office automation software as well suitable for parsing into VESPA-compatible tables for EPN-TAP (See Section 4.1).

A series of datasets from the previous Europlanet 2020-RI have been used as training datasets to establish the protocols.

The TA VUA office will act as contact providing advice if datasets fail to fully meet the FAIR principle. The VUA TA office will place the datasets on an open access server initially housed and supported at VUA.

A dedicate database is being established based at VUA using the national surfdrive storage option. The database will be open access and set up under the FAIR principle. Specific TA facilities are already integrated into specialised international databases. For example the data produced at the CSS facility in Grenoble:

All the spectroscopic data and metadata produced by the instruments of the CSS

facility (Grenoble, France) during the TA visits are stored in a dedicated 'CSS' database in the SSHADE database infrastructure (https://www.sshade.eu/db/css) where they can be easily searched and retrieved. SSHADE fully complies with the FAIR principles.

On the longer term (> 5 years) it has been agreed that the Europlanet Society will take the lead in administering long term data generated within the RI and other funded national and international projects.

5.1.3 Ethical aspects

There are no ethical aspects that need to be considered.

6 NA2

Work Package Beneficiaries and Partners	
IWF-OEAW	Space Research Institute, Austrian Academy of Sciences, Austria
UoE	University of Edinburgh, UK
UPV/EHU	University of the Basque Country
VU	Vilnius University, Lithuania
AMU	Adam Mickiewicz University, Poland
OBSPARIS	Observatoire de Paris, France

6.1.1 *Introduction*

The main objectives of the NA2 – “Coordination of Ground-based Observations” Work Package within the Europlanet 2024 Research Infrastructure (EPN2024-RI), as stated in the proposal, are:

- To coordinate a network of small telescope facilities (telescope diameters around 0.5-2.0m) to react fast and adequately to observational alerts;
- To coordinate professional and amateur long-term observational campaigns and timely constrained observations of Solar System objects, widening the participation of amateur astronomers in planetary science;
- To train and educate citizen scientists (amateur astronomers) and integrate them into the planetary science community;

- To assure that observational data of small telescope facilities is made available via the VESPA VA;
- To link this activity with all other WPs and to the diverse planetary sciences community in Europe and beyond.

To reach these goals, the so-called "Europlanet Telescope Network" (EPN-TN) was established which at the outset consisted out of 16 different telescope facilities in Europe and beyond, and now withwith 29 telescopes available in 17 observatories. Observers – professional and amateur observers – can apply for observations via the "Europlanet 2024 RI NA Call for Observations at the Europlanet Telescope Network" (call-website: https://bit.ly/2Br5LDt). The incoming proposals for observations will be reviewed by the NA2 Science Advisory Panel in an anonymized way on a bi-monthly basis. The observational data that will be produced through these observation campaigns or from the beneficiaries of NA2 will ultimately be made publicly available. To train and educate citizen scientists, NA2 will further organize and support amateur training workshops for which Workshop Guidelines will be prepared and made publicly available through the ~~recently~~ established website of NA2 (https://bit.ly/37SCiyj). The NA2 WP will finally develop an alert system for observations, which will notify and allow participating observatories and citizen scientists to select appropriate targets across the diverse range of planetary science topics. For this purpose, a dedicated website will be created, gathering easy-to-find information about observation campaigns and links to the (mostly external) tools for observation planning. This software/website will be developed and hosted by AMU and will be maintained on a GIT repository.

6.1.2 Data Summary

The data produced within NA2 can be categorized in four different types, i.e. observational data, personal data, software, and guidelines and documentation. These are described below in more detail.

6.1.2.1 Data types and formats:

1. Observational data:

   Observational data will either be produced by the beneficiaries (UPV/EHU, OBSPARIS) through observational campaigns, telescope facilities of the EPN-TN (see List of telescope facilities in the EPN-TN), or through external observers, who observe on one of the telescope facilities of the EPN-TN. The data will consist of standard formats for observational data such as images, movies, or maps (jpg, gif, tiff), or the FITS format, and are aimed to be made publicly available via EPN-TAP services linked to the VESPA VA.

2. Personal data:

   For evaluating proposals to the NA2 Call for Observation and for organizing amateur trainings and/or workshops for the organization of observation campaigns, NA2 will have to collect a restricted amount of personal data of natural persons from the applicants and workshop participants. These personal data will only be collected with the consent of the respective natural persons for

evaluating the proposals and organizing the workshops. NA2 thereby fully complies with the GDPR's privacy and security requirements.

3. Software:
   The observational alert system will be delivered in a form of a web service called PARSEC (PlAnetaRy SciencE Collaboration tool). The service will be deployed on a Linux server in AMU Astronomical Observatory Institute and is being developed in Python using Flask backend framework with Redis and RQ task queue, Bootstrap frontend framework and JavaScript. Some backend utilities are written in C/C++.

Guidelines, documentation, and workshops:

Guidelines for the organization of amateur training workshops will be established and made available online on the website of NA2 together with agendas, presentations, and minutes of these workshops. Any further documentation of observation campaigns, the telescope facilities, observational data, and/or of the observational alert system will be made publicly available through the website of NA2. Any document will be released in standard formats such as pdf, google docs, or google sheets. Workshops organized by NA2 will further be virtually accessible through tools such as Zoom or Go2Meeting.

6.1.3 FAIR data - findable, accessible, interoperable, re-usable

All observational data produced by the telescope facilities within observation campaigns supported by NA2 will ultimately be made open access based on the FAIR principle: free access and information retrieval. The observing teams will have a year following the supported observation nights to make use of new data before it is required to be stored and made open access. Longer periods are allowed for on-going PhD studies and industrial partners may further individually request a waiver based on the commercially sensitive nature of the data. This principle requires that all observing teams will store observational data they generate within supported observation nights on a server that itself is accessible to and searchable from the outside world. The respective publicly available server will ideally be accessible through an EPN-TAP service via the Virtual Observatory VESPA with its metadata being compliant with the VESPA metadata standards.
Should a telescope observing team not already have an open access data retrieval system they will be requested to upload it to a publicly available database that is ideally accessible through VESPA (such as PVOL, http://pvol2.ehu.eus/pvol2/), or to install a new EPN-TAP service that describes their observational data products and links it to VESPA. For setting up an EPN-TAP service, Europlanet 2024-RI can give technical consultations and provide related travel reimbursement if needed.
Intellectual property rights on the different observational data will be held by the respective observing teams. The beneficiaries of NA2, the observing teams, and the

telescope facilities thereby agree to implement and abide by the European Commission's Intellectual Property Code of Conduct.
The website of NA2 (https://bit.ly/37SCiyj) will further serve as the access point for all of the documentation of NA2 and will provide open access for the public to

- a description of NA2, its activities (including upcoming and past workshops), and the Europlanet Telescope Network;
- the NA2 Call for Observations;
- descriptions of the involved telescope facilities;
- the observational alert system;
- information on observation campaigns;
- workshop guidelines and documentation of the different workshops;
- other documents.

The observational alert system PARSEC will be free to use for any user. The source code will be available for anybody who wants to contribute to the project and aid its development via a git service. The software may become open source after the PARSEC web service is launched. If so, a GNU GPLv3 license will be used.
The documentation of NA2 and the observational alert system will be sustained after project-end based on best-effort and availability of resources.

7 References

Galluzzi, V., L. Guzzetta, L. Ferranti, G. Di Achille, D. A. Rothery & P. Palumbo (2016) Geology of the Victoria quadrangle (H02), Mercury, Journal of Maps, 12:sup1, 227-238, DOI: 10.1080/17445647.2016.1193777

Malin, M. C., J. F. Bell III, B. A. Cantor, M. A. Caplinger, W. M. Calvin, R. T. Clancy, K. S. Edgett, L. Edwards, R. M. Haberle, P. B. James, S. W. Lee, M. A. Ravine, P. C. Thomas, M. J. Wolff (2007), Context Camera Investigation on board the Mars Reconnaissance Orbiter, Journal of Geophysical Research, 112, E05S04, doi: 10.1029/2006JE002808.

McEwen, A. S., Eliason, E. M., Bergstrom, J. W., Bridges, N. T., Hansen, C. J., Delamere, W. A., ... & Kirk, R. L. (2007). Mars reconnaissance orbiter's high resolution imaging science experiment (HiRISE). *Journal of Geophysical Research: Planets*, *112*(E5).

Neukum, G., & Jaumann, R. (2004). HRSC: The high resolution stereo camera of Mars Express. In *Mars Express: The Scientific Payload* ESA Special Publication 1240, pp. 17-35).

Jaumann, R., Neukum, G., Behnke, T., Duxbury, T. C., Eichentopf, K., Flohrer, J., ... & Hoffmann, H. (2007). The high-resolution stereo camera (HRSC) experiment on Mars Express: Instrument aspects and experiment conduct from interplanetary cruise through the nominal mission. *Planetary and Space Science*, *55*(7-8), 928-952.

8 Annex A – List of VESPA datasets

8.1 Scientific datasets and Metadata catalogues

An up-to-date list of published *scientific datasets* and *metadata catalogues* is available from the VESPA portal .
DMPs for individual datasets are accessible from this page: https://voparis-wiki.obspm.fr/display/VES/Individual+DMP+of+EPN2024RI+collections

8.2 Documentation

The VESPA documentation is primarily managed on the project wiki. The VESPA Document Register provides links and metadata for all VESPA documentation produced on the wiki.
VESPA tutorials are listed on the VESPA web page (http://www.europlanet-vespa.eu/tutos.shtml), and are maintained on Github (https://github.com/epn-vespa/tutorials).
VESPA publications are listed in the VESPA wiki: EPN-2024-RI Publications.

8.3 Software

About License: open source is required for new developments. By default it is GPLv3. Case by case license selection (but still open source) for specific developments (update of existing tools). We do not list here the web portal softwares, which are not open source to prevent hacking hazards.
NB: the VESPA web portal code is available on request at support.vespa@obspm.fr, as described on the page Installing a local VESPA client.

9 Annex B - List of Recommended interoperable data formats and APIs

Format	Full Name	Metadata and configuration specifications	References
CDF	Common Data Format	ISTP, PDS4, EPNcore	http://cdf.gsfc.nasa.gov
FITS	Flexible Image Transport System	FITS, WCS, geoFITS (1)	http://fits.gsfc.nasa.gov
VOTable	Virtual Observatory Table	EPNcore	http://www.ivoa.net/documents/VOTable/
HDF5	Hierarchical Data Format 5	HDF5 1.6 EarthData	https://www.hdfgroup.org/HDF5/
netCDF	Network Common Data Format	netCDF CF	http://www.unidata.ucar.edu/software/netcdf/
geoJSON	geo-referenced JSON	IETF-RFC7946	http://geojson.org
geoTIFF	geo-referenced TIFF	GeoTIFF (webarchive.org)	https://trac.osgeo.org/geotiff
WMS	World Map Services	http://schemas.opengis.net/wms/	http://www.opengeospatial.org/standards/wms
HAPI	Heliophysics API	https://github.com/hapi-server/data-specification	https://github.com/hapi-server
das2	Das2stream	das2 ICD (v2.2.2)	https://das2.org

(1) Marmo et al. (2018) FITS Format for Planetary Surfaces: Definitions, Applications, and Best Practices. E&SS.doi:10.1029/2018EA000388

10 Annex C - VESPA beneficiaries preservation plan

Each Beneficiary Institute to put here their preservation plan: mainly, this is the physical infrastructure for your data storage.

10.1.1 CBK/PAN – Warsaw

The data and metadata of the CBK PAN databases, like LOFAR Jupiter observations, ionospheric data and space weather forecast messages, is stored in 7 physical machines running Debian 10. The json files database is working with the MongoDB database in version 3.4. Larger files are stored in NetCDF format in dedicated mass storage. The system is developed to keep at least one copy of data. The Mongo database allows us to keep high availability for over 99.9 % of json files, The large files database is maintained by the NFS shares, between physical machines, there is at least one copy of NetCDF file, the synchronization is running automatically by the rsync method every day.
The endpoint VESPA interface is ran by the VM managed by XenCenter 7.2. The snapshot of VMs is made manually every month, and can be restored on demand.

10.1.2 CNRS, OAS/CDS – Strasbourg

The software elements will be maintained by OAS/CDS as part of the Aladin software.
The Vizier_planets service relies on B/planets (https://vizier.cds.unistra.fr/viz-bin/cat/B/planets) which is one of the catalogues of the OAS/CDS VizieR service. It indexes a selection of VizieR catalogues related to the solar system and exoplanets. Its preservation follows the same rules as for all VizieR data (multi-site redundancy and backups) and satisfies the CoreTrustSeal requirements (https://www.coretrustseal.org)

The sixty planetary HiPS generated and maintained by the OAS/CDS are integrated into the CDS global storage and distribution system as HiPS node in accordance with the HiPS IVOA recommendation. Physically, these data, like all HiPS, are duplicated on the two disk bays of 1.6Pb effective storage localized on the two geographical computing sites managed by the CDS; external mirrors also exist, in particular at PADC / Observatory of Paris.

10.1.3 CNRS, IPAG – Grenoble

The data and metadata of the SSHADE solid spectroscopy database infrastructure will be stored at the OSUG DataCenter of the Observatoire des Sciences de l'Univers de Grenoble.
The data is stored on the OSUG Datacenter using the UGA's massive storage project SUMMER based on NetApp technology. This storage space is duplicated on three distant sites on the university campus. All data stored is backed-up and replicated

with the following specifications: up to 255 snapshots per volume, asynchronous replication in a 1-hour delay on another site, daily backup up to 30 days on a distant site. All storage use RAID-DP to ensure a two-disk fault tolerance, with minimal performance impact. Every technical equipment (disk controller, alimentation, network link, safety disks) is redundant on-site. The technical implementation of this service allows an availability of 99.99 percent.
The servers hosting the webservice are hosted on the mutualised VMW are platform of the OSUG Datacenter. This platform is composed of 2 x 2 X670 servers on distant sites linked by dedicated optic fibers. The VM are replicated on both sites, and the migration of operations can be done using vMotion to allow seamless transition in case of server failure. The VM use regular snapshot that can be restored on demand.

10.1.4 CNRS, IPSL – Paris

The VO services and data are hosted on dedicated Virtual Machines (VM) on IPSL's mutualized VMcenter at ESPRI. Each service has a companion VM "sandbox" for implementation of developments and tests. All of IPSL's VMs are backed up on a nightly basis.

10.1.5 CNRS, IRAP – Toulouse

The data and metadata of the CDPP/AMDA database are stored on the CDPP server (Linux CentOS) and saved by the IRAP Data Center and by the "Service de Sauvegarde et d'Archivage de Données" of the Observatoire Midi-Pyrénées (OMP). This service is responsible for the preservation of all data managed by institutes depending on the OMP. Its infrastructure is based on three rooms (Toulouse and Tarbes) systems, an efficient and secure network, a virtual platform and a storage and backup of more than 300 TB capacity.

Other software elements are maintained by IRAP as part of the CASSIS software.

10.1.6 DLR – Berlin

DLR Planetary Research Institute has in-house and remote server managed by our IT provider. The disks containing the Data are under versioned backup every 12 hours and on RAID-6 system disks. The raw the NASA MESSENGER DATA are preserved on NASA PDS Imaging Node Mission Page, and duplicated between several physical machines. The NASA physical infrastructure is outside the scope of this document.
The data+metadata are derived from raw PDS3 MESSENGER/MASCS and are ingested in Postgres DB on RedHat8 server. This server is entirely dedicated to the data storage.
We execute also regular manual backup to text json format and SQL dump.
The VESPA related files are stored in DLR internal Github server and will be mirrored on Github.

10.1.7 UPV/EHU – Bilbao

Amateur observations of solar system planets will be stored in a devoted facility. UPV/EHU hosts a web service currently named PVOL (Planetary Virtual Observatory Laboratory, http://pvol.ehu.eus) devoted to amateur observations of Giant Planets. This service is continually maintained and will be upgraded to contain observations of other planets (Venus and Mars). The infrastructure consists of a webserver and a regular backup of the data. The service stores and facilitates access to amateur observations of giant planets over the last decade.

10.1.8 IASB-BIRA – Brussels

IASB has an important IT-infrastructure to support all the research and operational activities. The parts of the infrastructure which are of most interest to the proposal are the storage infrastructure, the HPC compute clusters, the interconnecting network and the internet connectivity. Additionally, the ICT team has dedicated personnel for the support of these systems and their users. The ICT infrastructure of the institute has sufficient capacity and resources available for the proposed tasks in this proposal.
The current storage facilities are composed of a highly available high-performance NAS (Network attached storage) server and a dedicated HSM (hierarchical storage management) storage system. The NAS servers are a cluster of Netapp FAS-3240 servers with an additional FAS-3210 backup node. The total available storage capacity on the Netapp filers is +/- 200 TB. The HSM system is a hybrid disk storage/tape storage system with automatic migration of files between disk and tape. This system currently has a capacity of 200TB on tape and 60TB on disk. The server can be easily extended in capacity for future needs by adding tape storage. Both of these storage systems are accessible from all compute servers as well as from the individual user workstations. Full backups are taken of all data on both storage systems. To assist the users in the management of their data and to provide professional management of the storage infrastructure the institute has a dedicated 'data manager'. This member of the ICT team oversees all data storage and distribution tasks and optimises the use of the infrastructure. He also serves as the central contact point for all users concerning data related questions: data import, storage, backup, distribution.

10.1.9 Jacobs University – Bremen

Data hosted on PlanetServer (Planetary Science Data Service of http://earthserver.eu, e-infra project #654367) are maintained by Jacobs University and are going to be interfaced with the VESPA database.
Data, of several tens of TB, are available on redundant RAID systems. Platforms are linux-based (CentOS). During the lifetime of the projects (and any eventual follow-up) data will be maintained, but no long-term storage and preservation is guaranteed. Nevertheless, all data processing and ingestion routines, as well as server and client software are available and so will stay beyond the lifetime of the project(s), in order to allow reproducibility. Raw data used are derived from PDS and PSA, thus, long-term available.

Software of PlanetServer developed solely within the EarthServer project is available on https://github.com/planetserver. Any other additional software developed within EuroPlanet-H2020-RI Vespa on the VESPA GIT repository. In both cases, the availability of code and related documentation is not bound to project duration, thus longer-term.

10.1.10 OATS/INFA – Trieste

Data services will be provided by INAF - the Astronomical Observatory of Trieste with the support of Data Center of IA2 – INAF. Currently the local VESPA server (already available at http://vespa.ia2.inaf.it/) is housed at IA2 as a virtual server (VMware based, DMZ) which can be upgraded according to the data services requirements. The storage system is managed by the IA2 data center including preservation and data base management (PostgreSQL). It will host the data services, which are not expected to be highly demanding in terms of resources (~ few tens of TBs). The data will be kept online (RAID file system) and preserved on tape library, the virtual server itself is backed up weekly (in virtual machine snapshots). The server runs DaCHS 2.5 on the Debian 11.1 operating system. The server is already configured to preserve configuration and metadata of the data services using the VESPA project Gitlab instance. Both storage and data base are monitored through the IA2 internal NAGIOS System, the DaCHS Server itself can be added to this monitoring system.

10.1.11 OBSPARIS – Paris/Meudon

The data preservation policy of the Observatory of Paris relies on its data centre PADC (Paris Astronomical Data Centre). The access to services is enabled by virtual servers that can be easily reconstructed and that are independent of the physical infrastructure hosting them. Operating systems hosting the services are FreeBSD or Linux Debian, which are open source. Their configuration is managed using Puppet, which allows to automatically install applications, monitor and backup systems, and ensures straightforward recover plans. The data are stored and replicated on two physical sites distant by more than 7 km and independently connected to the internet (Paris and Meudon sites of the Observatory of Paris). The OBSPARIS services will be eventually served from both sites to ensure a high availability. The data are stored on a ZFS file system. This solution enforces data integrity. It is also replicated by block, as required by the large number (several millions) of files hosted at PADC. It also uses a differential snapshot capability to synchronize the two archives. In addition to the two instances of disk storage area, PADC has a system on tapes using 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 catalogue. The 1.3 PBytes robotic tape system of OBSPARIS ensure the backup and preservation of the project data. PADC hosts and maintains a gitlab used to store service files from all beneficiaries. This system is accessible to all contributors through EduTEAMS AAI (VESPA virtual organisation). It is used as a backup facility to ensure sustainability, but will also allow deploying the services on EOSC in the future.

The sustainability of PADC is ensured by permanent civil servant positions at OBSPARIS, which has kept its duties and buildings since 1667.

10.1.12 OeAW – Graz

All data at the space research institute in Graz (IWF-OeAW) are stored on two mirrored RAID systems that are distributed over two server rooms. The capacity of the storage system is roughly 600Tb. The file-systems used are Solaris 11-ZFS, Linux EXT4 and Windows NTFS. The data is being backed up on a daily basis, a full backup is performed once a week. The backup hardware consists of a HP solution, using tape as the storage medium. The backup software used is Legato Networker. The platforms for operating system virtualization are XenServer and VMWare, the oprating systems used include Windows 2008, Windows 2012 R2, CentOS 6,7, Ubuntu14.04 and Debian 8. Xitrix CenApp is providing application-virtualization of Matlab, Mathematica, Comsol IDL etc. Monitoring of the system is provided via ZABBIX and NAGIOS. The institute also hosts a small HPC cluster with approximately 700 cores. The infrastructure is permanently maintained by 4 full-time employees in Graz, basic IT services as e.g. Email, DNS and network hardware are provided via the Vienna based OEAW/ARZ.

10.1.13 SINP-MSU – Moscow

TBC – collaboration suspended due to Ukrainian invasion by Russia

10.1.14 SpaceFrog – Toulouse

Will not provide data services or software

10.1.15 UCL – London

TBC

10.1.16 University of Bristol – Bristol

The software elements will be maintained by University of Bristol as part of the TOPCAT software. This software is managed using the Git/GitHub distributed version control system, which provides a high level of resilience.

10.1.17 University of Heidelberg – Heidelberg

The software elements will be maintained by Univ of Heidelberg as part of the DaCHS software.

Data services are maintained together with other VO services in astronomy, as part of the gavo data centre.

11 Annex D - GMAP file formats used and envisaged

File formats used for day-to-day activities and sharing, archiving might not be the same. PLANMAP envisages formats for both use and sharing, and archiving will be produced and these are indicated below. When relevant, appropriate PDS (PDS4) labels will be added to describe the data products, depending on actual arrangements with ESA on data delivery. Data shared via VESPA will not bear any PDS label and will contain metadata as needed for operating VESPA services and accessing and using data with commonly used geospatial software (e.g. GDAL/OGR-compatible).

11.1 Base formats depending on data type

Dataset type	Format	Common extensions	Products	Notes
Raster	GeoTiff with appropriate type and signedness (compressed or not)	.gtiff, .tif, .tiff	Imagery, possibly multi-band imagery and Digital Terrain Models	
	jpeg2000 (lossy or lossless)	.jp2 .jk2	Imagery for basemaps/mosaic for which higher compression rates is desirable	
	Portable network graphics	.png	Imagery for web publication and GIS, imagery intended only for visualization (i.e. grayscale/colour mosaics)	
	ASCII Gridded formats supported by gdal (arc/Info, GRASS, etc)	any [e.g. .txt, .grid, etc]	alternative to geotiff for Digital Terrain Models	geotiff is still preferable (e.g. no round-off errors)

	World (georeferencing) files	.twf, .pwf, .jwf (or similar)	Gereferencing information for raster dataset which is coupled with each raster file	
Vectors	Geopackage (OGC)	.gpkg	All mapping products in vector format	
	Shapefile	.shp (.dbf, .shx , .sbn, .prj)		
	Geographical markups: JSON (GeoJSON)/XML (GML)	.json, .xml	Georeferenced products shareable on the web, WFS-served dataset	
	Relational DB (PostgreSql + PostGIS / SpatiaLite)	----	Dataset meant to be served by Web Feature Services [WFSs]	Internal use only
	Styled Layer Descriptors	.xml	XML files describing the styles to be used when displaying vector layers	
3D	Stanford Triangle Format, Wavefront OBJ, Stereolithography an d comparable	.ply, .obj, .stl	Three dimensional meshes. Geological meaning must be provided by metadata.	Neutral file formats must be preferred. Might be equipped with texture files (i.e. for terrain with orthophoto)

	VTK File Formats	.vtu, .vtp, .vtk	Commodity format for data exchange. Useful for products that require the preservation of scalar fields associated to triangles or vertices of the meshes and scientific analysis via VTK.	
	Geological Aware Formats	application-specific	3D geological models from 3D geomodeling software packages	Neutral file formats must be preferred
Hierarchical archives	Planteary Data System (v3,v4)	.pds, .img	Archived dataset with variable content	