EPN2020-RI

EUROPLANET2020 Research Infrastructure

H2020-INFRAIA-2014-2015

Grant agreement no: 654208

Document: VESPA-026-PL v2.1

VESPA Data Management Plan

Date: 2019-08-29

Start date of project: 01 September 2015

Duration: 48 Months

Responsible WP Leader: OBSPARIS / Stéphane Erard

Project co-funded by the *European Union's Horizon 2020 research and innovation programme*		
Dissemination level		
PU	Public	•
PP	Restricted to other programme participants (including the Commission Service)	•
RE	Restricted to a group specified by the consortium (including the Commission Services)	•
CO	Confidential, only for members of the consortium (excluding the Commission Services)	•

Project Number	654208
Project Title	EPN2020 - RI
Project Duration	48 months: 01 September 2015 – 30 August 2019

Abstract: VESPA is an virtual research infrastructure. It is fed by contributors, who are sharing their science datasets using the VESPA infrastructure. This document presents the Data Management Plan for VESPA, including science dataset, metadata catalogues, documentation and software.

Document history (to be deleted before submission to Commission)

Date	Version	Editor	Change	Status
30 nov. 2015	0.1	Baptiste Cecconi	Initial Release	DRAFT
1 déc. 2015	0.2	Baptiste Cecconi, Manuel Scherf, Tarek Al-Ubaidi	Update after telecon with IWF/Graz	DRAFT
2 déc. 2015	0.3	Baptiste Cecconi	Added DMP for documents and software	DRAFT
3 déc. 2015	0.4	Baptiste Cecconi, Vincent Génot, Aurélie Kasprzak, Christiane Adam	Many corrections and updates. Added more details for software.	DRAFT

3 déc. 2015	0.5	Baptiste Cecconi, Ann-Carine Vandaele	Added IASB-BIRA inputs Added existing EPNcore metadata services.	DRAFT
7 déc. 2015	0.6	Baptiste Cecconi, Stéphane Erard, Christiane Adam, Francoise Genova, Chiara Marmo, Bernard Schmitt, Anni Määttänen, Damien Albert	Included various inputs Updated datasets names Simplified document structure	DRAFT
7 déc. 2015	0.7	Baptiste Cecconi	First export and draft delivery to OU	DRAFT
8 déc. 2015	0.8	Baptiste Cecconi, Chiara Marmo, Damien Albert, Aurélie Kasprzak	Adding Preservation plan for CNRS-GEOPS Update of Figure 1 Small wording changes	DRAFT
9 déc. 2015	0.9	David Pisa, Gangloff Michel, Baptiste Cecconi, Ricardo Hueso	Update IRAP, IAP, UCL, EHU, ObsParis contributions	DRAFT
14 déc. 2015	0.10	William Thuillot, Josselin Desmars, Daniel Hestroffer, David Bancelin	Adding OBSPARIS-IMCCE inputs	DRAFT
20 janv. 2016	0.11	Baptiste Cecconi, Chiara Marmo, Angelo Pio Rossi	Finalizing DMP for submission to EPN coordination	DRAFT
21 janv. 2016	0.12	Baptiste Cecconi, Maria Teresa Capria, Ricardo Hueso	Adding IAPS/INAF and UPV/EHU inputs	SUBMITTED
10 févr. 2016	0.13	Baptiste Cecconi, Stéphane Erard	Including comments from OU	DRAFT
14 févr. 2016	1.0	Baptiste Cecconi	Final changes	ISSUED
3 mai 2017	1.1	Baptiste Cecconi	Fixed typos and Appendices	DRAFT

22 août 2017	2.0	Baptiste Cecconi	version 2.0 changes	ISSUED
19 août 2019	2.1	Baptiste Cecconi	Update lists of products	ISSUED

Table of Contents

Acronyms

- DMP – Data Management Plan
- EPN2020RI – Europlanet-H2020 Research Infrastructure
- EPNcore – Europlanet Core Data Model
- FTP – File Transfer Protocol
- HTTP – Hyper Text Transfer Protocol
- IPDA – International Planetary Data Alliance
- IVOA – International Virtual Observatory Alliance
- NA – Networking Activity work package
- OGC – Open Geospatial Consortium
- PADC – Paris Astronomical Data Center
- PDS – NASA Planetary Data System
- PSA – ESA Planetary Science Archive
- RDA – Research Data Alliance
- SPASE – Space Physics Archive Search and Extract
- TA – Transnational Access work package
- TAP – Table Access Protocol
- URL – Uniform Resource Locator
- VA – Virtual Access work package
- VESPA – Virtual European Solar and Planetary Access
- W3C – World Wide Web Consortium

Beneficiaries

- OeAW – IWF, Graz – Austria.
- IASB-BIRA – Brussels – Belgium.
- IAP – Prague – Czech Republic.
- Jacobs University – Bremen – Germany.
- UPV/EHU – Bilbao – Spain.
- OBSPARIS – IMCCE, Paris; LESIA, Meudon – France
- CNRS – IRAP, Toulouse; GEOPS, Orsay; IPAG, Grenoble; LATMOS, Paris; OAS/CDS, Strasbourg – France
- GFI Informatique – Toulouse – France
- IAPS/INAF – Rome – Italy
- IGS/PAS – Wroclaw – Poland
- UCL – London – United Kingdom
- Uni Bern – Bern – Switzerland

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 EPN2020RI beneficiaries from TA, NA and VA work packages, as well as any external team willing to share data using VESPA. There is thus two types of dataset shared in VESPA: the *scientific datasets*, which are the result of a scientific observation or analysis produced by each contributor; the *metadata catalogues*, which are derived by *VESPA contributors* from *scientific dataset* and are shared in a common format (EPNcore[1]). 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. Each EPN2020RI beneficiary contributor has to contribute to the EPN2020RI DMP.

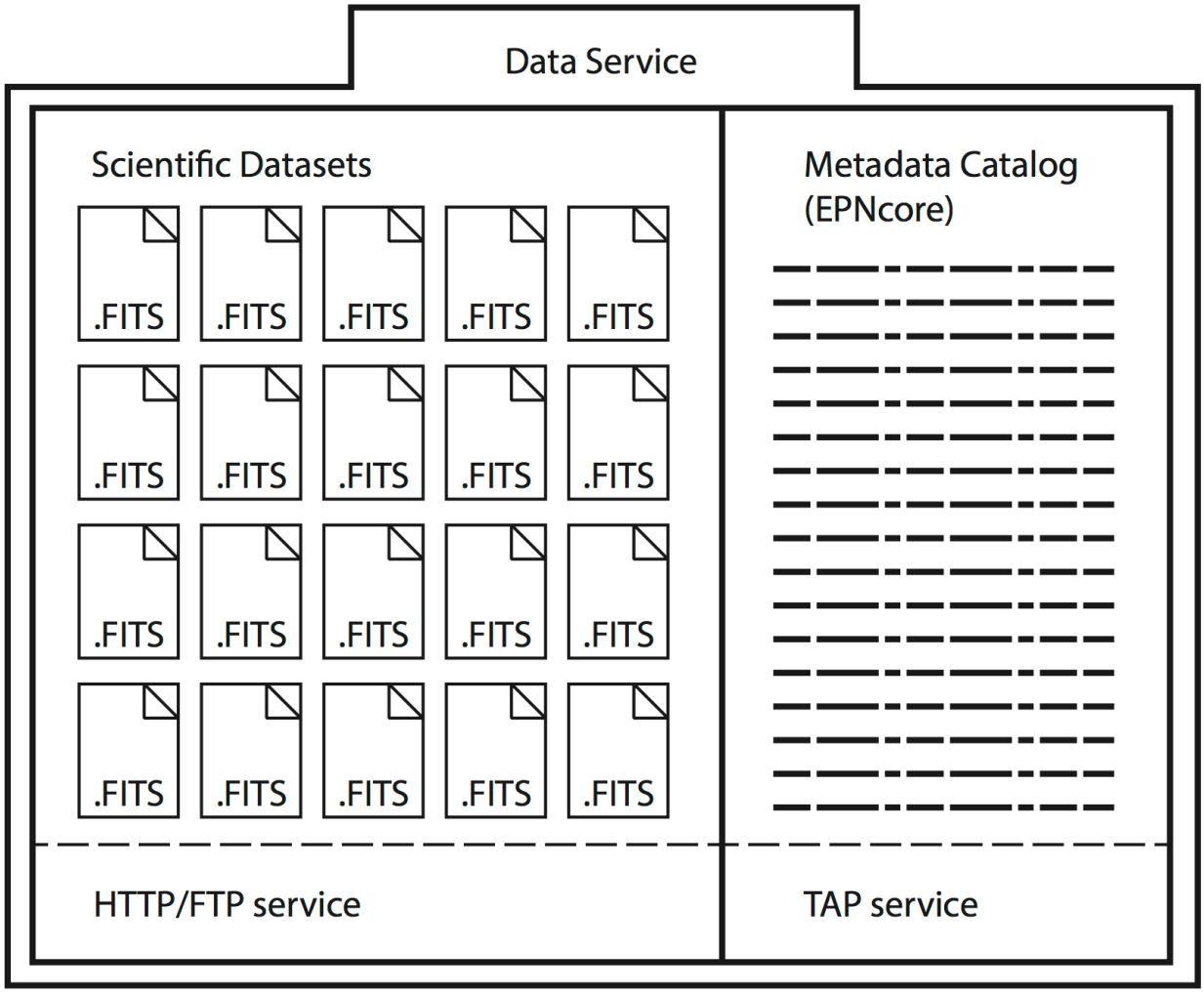

Figure 1. The contributors are generating *scientific datasets*, using standard file formats (here FITS files) and associated *metadata catalogues* (with the EPNcore specification). The data

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.*

Data types, formats, standards and capture methods

Scientific datasets

The *scientific datasets* generated by the EPN2020RI beneficiaries of the VESPA work packages are provided on the 1st table of 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.

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.* We expect at least 40 *metadata catalogues* at the end of the project (12 *metadata catalogues* are accessible on day 1). All EPN2020RI 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 project. In the course of the Europlanet-2020-RI 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 TAP.

Documentation

VESPA *documentation* consists of various types of documents:

- Assessment studies are working documents discussing the future developments or standard adoptions.
- Reports are summarizing activities.
- Recommendations propose good practices to VESPA contributors, and are submitted for endorsement/adoption as standards to international consortia (IAU, IVOA, IPDA).
- Tutorials contains 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 are 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 EPN2020RI quality guidelines.

Documentation is prepared using a document template in the VESPA discussion tool. 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).

Software

VESPA *software* are tools and libraries. They are developed to enhance the access to *scientific datasets* and *metadata catalogues*. VESPA-JRA tasks 3 to 8 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* are developed using a GIT repository, enabling versioning and backup of codes. Good coding practices (such as automated testing procedures, commenting...) are recommended. Open licences improve maintainability. Modularity and frequent comments improve code understandability for future developers.

VESPA *software* are developed in many languages (Python, Java, Javascript...).

Data exploitation, sharing and accessibility

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 access 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.

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 licence, with a reference to a document indicating the rules of use (citation and acknowledgement policy).

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 EPN2020RI 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.

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.

Software

All software developed in the frame of VESPA activity are open source. The GPLv3 license is strongly recommended for codes developed in the projet. 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.

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.

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).

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 are 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 testing interface 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 (WP6 task 4) 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* can not commit to long term preservation of their datasets, usage of community-based preservation platforms such as Zenodo (or alike) is recommended.

Annex A – List of VESPA datasets

Scientific datasets

The list of scientific datasets is available in Table 6.1 of Deliverable 6.14 (4th VESPA Annual Report).

Metadata catalogues

All metadata catalogues are formatted according to EPNcore specification. An up-to-date list of services is available from the VESPA EPN-TAP Services page.

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: Publications

Other documentation produced locally by VESPA contributors are mentioned below:

Document Name	Description	Institute	Access	URL or document
ASIMUT User Manual	Complete description of the ASIMUT code	IASB-BIRA	open	http://planetary.aeronomie.be/en/projects.htm#proj_asimut

Document Name	Description	Institute	Access	URL or document
SSDM datamodel	Solid Spectroscopy Data Model of the SSHADE database	IPAG / OSUG-DC	open	https://wiki.sshade.eu/sshade/documentation/ssdm
SSHADE database managers tutorial	series of presentation describing how to prepare and upload data files	IPAG / OSUG-DC	restricted to SSHADE database managers	https://wiki.sshade.eu/sshade/provider
SSHADE database users tutorial	series of presentation describing how to prepare and upload data files	IPAG / OSUG-DC	open	https://wiki.sshade.eu/sshade/interface
Giant Planet Magnetopheres User Guide	User guide for data product of the Giant Planet Magnetopheres Dataset, including references to papers	UCL	open	https://zenodo.org/record/3353721
Description of the Asteroid orbital data base	Reports on the source of the data, on the process to set up the database and description of the content	OBSPARIS-IMCCE	open	DynastVO-draft.pdf
3DView user manual	User documentation of whole 3DView tool with latest evolutions.	GFI Informatique	open	http://3dview.irap.omp.eu/other/3DVIEW_Users_Guide_2_0.pdf

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.

Software Name	Repository	Description	Institute	Language	License
VOTable2GeoJson	https://github.com/epn-vespa/VOTable2GeoJson	Convert VOTables into GeoJson	OBSPARIS	Python	GPLv3
CESIUM Planetary App	https://github.com/epn-vespa/cesium	Planetary visualization tool using CESIUM	GEOPS	JS	Apache 2.0
GDAL (FITS compatibility)	https://github.com/epn-vespa/gdal	Complete FITS implementation for planetary surfaces	CNRS-GEOPS	C / C++	X/MIT license
ALADIN	http://aladin.u-strasbg.fr/aladin.gml	Addition of relevant capacities discussed with the project in the Aladin software. Initial proposal: adaptation/extension of HiPS for planetary sciences and update of the ALADIN client software accordingly.	CNRS - OAS/CDS	Java	GPLv3

Software Name	Repository	Description	Institute	Language	License
LecturePDS	https://github.com/epn-vespa/LecturePDS	PDS3 reader for IDL/GDL	OBSPARIS	IDL, GDL	BSD-like
IDL_VOtable	https://github.com/epn-vespa/IDL_VOtable	VOtable reader and writer for IDL / GDL	OBSPARIS	IDL, GDL	GPLv3
3Dview	https://gitlab.irap.omp.eu/CDPP/3DVIEW/tree/master	Adding Spice Kernels for field of views, EPN-TAP query and visualisation through SAMP, and surface mapping in 3Dview (http://3dview.cdpp.eu)	GFI informatique	Fortran,C,Ksh,Java	GPLv3
PlanetServer client (server) code	https://github.com/planetserver	web client and server script components (rasdaman community edition GPL and separately avaiable - http://rasdaman.org)	JacobsUni	JavaScript, Python	GPLv3
EPN-TAP library	https://gitlab.irap.omp.eu/OV-GSO-DC/EpnTAPClient	EPN-TAP client to be used in Java-based tools. Currently implemented in CASSIS and 3Dview.	CNRS-IRAP, GFI	Java, Javascript	GPLv3
MASER	https://github.com/maserlib/maser4py	Set of python modules for low frequency radioastronomy	OBSPARIS	Python	GPLv3
VESPA.app	https://github.com/epn-vespa/vespaapp	Mobile vision of the VESPA portal, dedicated to imaging data products	JacobsUni	javascript	GPLv3

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/

Format	Full Name	Metadata and configuration specifications	References
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

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.

CNRS – OAS/CDS

The software elements will be maintained by OAS/CDS as part of the Aladin software.

CNRS – GEOPS

Software developed at GEOPS emanate from open-source, stable community-based software (eg GDAL library). Modifications will be validated with the official team and merged with the projects.

The data and metadata of the FRIPON project will be archived at and distributed by the Integrated Data and Operation Center (IDOC) at OSU Paris Sud. IDOC is responsible for the distribution of a number of spatial mission data products.

CNRS – IPAG

The data and metadata of the SSHADE solid spectroscopy database infrastructure will be stored at the OSUG DataCenter of the Observatory 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 allow an availability of 99.99 percent.

The servers hosting the webservice are hosted on the mutualised VMWare 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.

CNRS – IRAP

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 .

CNRS – LATMOS

The IT-infrastructure of the LATMOS is based on 2 data centres. One facility is based on the Observatoire Versailles Saint-Quentin (OVSQ) - Guyancourt and the second on Université Pierre et Marie Curie (UPMC) - Paris. These facilities are secure with access control and with power and air conditioning redundancy. All systems are managed by the IT team of the laboratory that ensures their performance and security (network, routing, access control, storage). The infrastructure consists of access networks 1Gb/s redundancy, a virtualization cluster (VMWARE), a secure storage of 200To (SAN), a filtering system, access control and a backup system (ASG TimeNavigator). The system will be hosted on these platforms to ensure efficiency and security.

UPV/EHU

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.

GFI Informatique

Not applicable. CNRS/IRAP will preserve and curate the data produced by GFI Informatique.

IAPS/INAF

The scientific datasets and the corresponding metadata will be stored in a devoted HW facility. IAPS is hosting a Data Center devoted to the instrument operations and the management of the data acquired in the space missions in which the Planetology Department is involved. The facility is continually maintained and upgraded, with adequate backup facilities.

MATISSE is hosted in a similar, bigger facility maintained by the Italian Space Agency (ASI).

IASB-BIRA

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 manger’. 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.

IAP

The interface for a multidimensional wave analysis (iPECMAN) will be maintained by IAP and interfaced with VESPA database. The software will be stored at the VESPA GIT repository. The optional standalone web application might be operated at IAP. For this purpose, a dedicated

server running on CentOS operating system with a sufficient performance and software is available. For temporary user data, a storage with a total capacity of 60 TB can be used. Long-term data storage is not planned at IAP.

IGS/PAS

The data of IGS/PAS will be stored and maintained in the SSHADE database infrastructure at IPAG / OSUG-DC.

Jacobs University

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 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.

OeAW

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.

OBSPARIS

The data are stored and replicated on two physical sites distant by more than 7 km and independently connected to the internet. The OBSPARIS services will be eventually served from

both sites to ensure a high availability. The data preservation policy relies on PADC. 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, monitoring and backup systems, and ensure straightforward recover plans. 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 is also using a differential snapshot capability to synchronize the two archives. 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 catalogue. The 1.3 PBytes robotic tape system of OBSPARIS ensure the backup and preservation of the project data. The sustainability of PADC is ensured by permanent civil servant positions at OBSPARIS, which has kept its duties and buildings since 1667.

UCL

Data and metadata for the UCL-based service are currently stored on a local server machine, 'astroweb', which is also used to host web services for other projects.

At present, as part of the ongoing software development, the CVS tool is being used to manage the development of the files necessary to run the UCL-VESPA service. This 'development repository' exists on a separate and secure local machine.

In terms of future development and preservation of the software, we have thus far used:

- Zenodo provides a doi for the Magnetodisc Field Modelling and Particle Tracing Tutorial (doi:10.5281/zenodo.3353720) with permanent link https://zenodo.org/record/3353721.
- The VESPA github service (https://github.com/epn-vespa/tutorials) is mirroring the MDISC zenodo archive.