



# eur PLANET 2024

Research Infrastructure

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 D1.8

**Deliverable Title:** VAs 2nd year External Board Review  
**Due date of deliverable:** 31/07/2022  
**Nature<sup>1</sup>:** R  
**Dissemination level<sup>2</sup>:** PU  
**Work package:** WP01  
**Lead beneficiary:** JacobsUni  
**Contributing beneficiaries:**  
**Document status:** Final

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

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)

# Euromlanet Virtual Access Review Board 2021 Report

## [1. Reference period](#)

## [2. Composition of the review board](#)

## [3. General Comments](#)

### [3.1 Objectives](#)

### [3.2 Impact](#)

### [3.3 Dissemination](#)

## [4. Specific per-VA and Data Management Plan Comments](#)

### [4.1 VESPA](#)

#### [4.1.1 Highlights](#)

#### [4.1.2 Specific issues](#)

#### [4.1.3 Specific actions suggested](#)

### [4.2 SPIDER](#)

#### [4.2.1 Highlights](#)

#### [4.2.2 Specific issues](#)

#### [4.2.3 Specific actions suggested](#)

### [4.3 GMAP](#)

#### [4.3.1 Highlights](#)

#### [4.3.2 Specific issues](#)

#### [4.3.3 Specific actions suggested](#)

### [4.4 ML](#)

#### [4.4.1 Highlights](#)

#### [4.4.2 Specific issues](#)

#### [4.4.3 Specific actions suggested](#)

### [4.5 Data Management Plan](#)

#### [4.5.1 Highlights](#)

#### [4.5.2 Specific issues](#)

##### [VESPA](#)

##### [SPIDER](#)

##### [GMAP](#)

##### [ML](#)

#### [4.5.3 Specific actions suggested](#)

## [5. Any Additional comments and suggestions](#)

# 1. Reference period

The reference period for this report is 12 months, from February 2021 through January 2022.

# 2. Composition of the review board

Anne C. Raugh, PDS Small Bodies Node, University of Maryland, USA

Christophe Arviset, ESA, Madrid, Spain

Chiara Marmo, Paris Saclay University, France

Beatriz Sanchez-Cano, University of Leicester, UK

# 3. General Comments

In general, the VA projects have done well in mitigating and managing the delays imposed by the pandemic and the related work and travel restrictions. Schedules have been adjusted, virtual alternatives to face-to-face meetings have been employed, and the overall impacts to the planned work packages have been minimised. All teams are to be commended for their extraordinary efforts during this period.

Excellent progress has also been made across the board in creating a unified and welcoming web presence consistent with the Europlanet branding. The individual portals provide the customised experience users recognize, appropriate to each of the projects.

Overall, the great majority of the issues raised in the previous report have been addressed for the better. It has been a productive year.

## 3.1 Objectives

**Has the set of VA activities met the objectives in the relevant period as described in the Description of Action? If not please provide suggestions. If not applicable to the current reporting period, please state it.**

The objectives set, in some cases revised in response to pandemic impacts, have been met, and in some notable cases exceeded. Deliverables were produced and received in good time; planned activities took place; and the milestone schedules, which have been reviewed and updated, look reasonable and achievable.

## 3.2 Impact

**Has the set of VA activities met the expected impact in the relevant period as described in the Description of Action? If not please provide suggestions. If not applicable to the current reporting period, please state it.**

Each project has presented evidence of impact in terms appropriate for each. Those projects offering direct services through their web portals provided access metrics that indicate users are finding and using the services in increasing numbers; abstracts for papers presented at conferences are increasing despite the lack of in-person conferences; and some citations of services were noted despite the general lack of support for citing data products still found in large parts of the publishing community. Community response to workshops has been positive and resulted in new providers, products, and services becoming available.

## 3.3 Dissemination

**Has the set of VA activities disseminated and exploited results in the relevant period as described in the Description of Action? If not please provide suggestions. If not applicable to the current reporting period, please state it. If not applicable to the current reporting period, please state it.**

Although still somewhat handicapped by the ongoing limitations on in-person gatherings around the world, dissemination tasks involving personal interaction through presentations at conferences and workshops have been successful in both virtual and hybrid forms. Publication counts are increasing, as are citations and the number of users accessing services. There has been success in increasing dissemination across all projects.

# 4. Specific per-VA and Data Management Plan Comments

## 4.1 VESPA

As the dissemination portal for data produced by other VA projects, the success of VESPA is key to the success of the VA work packages as a group. VESPA has made good progress, and is even ahead of schedule on some milestones.

### 4.1.1 Highlights

A combined workshop was held in late 2021 to replace the pandemic-delayed 2020 and 2021 scheduled workshops. 35 participants attended; 4 projects were selected prior to the workshop; 4 more were proposed during the workshop; and 3 other services were published through external collaborations.

VESPA beneficiaries published five new services; updated or upgraded nine additional services; and drafted six new services planned as part of the VESPA VA work packages. The SSHADE service, in particular, saw the addition of 50 new datasets (>1000 spectra) as well as significant updates to the user interface and dynamic plotting tool, and virtual meetings and training sessions with 16 partners (8 of them new) to add Raman band lists and bands of solids to the collection.

The backlog in bringing amateur observations online is being addressed, with 57.000 amateur images incorporated into the PVOL service; assessment of spectral data for inclusion is ongoing.

VESPA has applied two systems for providing DOIs for managed data: one internal to SSHADE and one for more general use.

VESPA is also having significant impact in the community through its work developing and disseminating standards through the International Virtual Observatory (IVOA); its participation in the NASA Planetary Data Ecosystem Review Board; as the Europlanet representative to the International Planetary Data Alliance (IPDA); and its participation and organisation of dedicated sessions in the Europlanet Science Conference (EPSC) and American Geophysical Union (AGU) Conference.

The VESPA section of the DMP is very clear, with well identified data types and details of the associated responsibilities for each one of the VESPA curators.

#### 4.1.2 Specific issues

None noted.

#### 4.1.3 Specific actions suggested

The access metrics for VESPA are useful. It would be helpful to see them presented as visualisations as well as raw numbers. For example, it would be interesting to see trends in both publication and use of services being developed as part of Europlanet 2024.

As a special case, the amateur data being presented in PVOL is of particular interest. It would also be useful to see metrics and visualisations for the amateur services as they become available, both for data access and for breakdowns of how many individuals or collaborations are contributing what kinds of data.

## 4.2 SPIDER

The SPIDER project had some significant successes for this review period, including a much improved portal, involvement in several planned and active missions, and a number of publications.

### 4.2.1 Highlights

The SPIDER portal shows significant improvement over the last review period, as do a number of the databases and tools provided through that portal. The access metrics for the databases show that they are seeing regular traffic.

A Planetary Space Weather session at EPSC 2021 (Session T28: *Planetary space weather and space weathering on airless bodies*) drew 12 oral and poster submissions.

SPIDER services were used during BepiColombo's Mercury and Venus flybys, are being used for Solar Orbiter, and will be used for planning future BepiColombo phases as well as for the JUICE mission and other missions and collaborations. The existing interactions have already resulted in several publications.

One new space weather service has been prototyped, and a new run-on-demand service is well into development.

### 4.2.2 Specific issues

While there is significant improvement in the SPIDER portal, a couple of issues still remain with the legacy Planetary Space Weather Services (PSWS) site:

- The PSWS is described as being completed, but some of its services are still described in the future tense ("will develop", e.g.).
- The PSWS site still contains non-working links. For example, on the Prediction page, the link to TAILCATCHER, returns a "Page not found" message (and is described as something that "will be developed"), and the Meteor Showers link to VESPA returns an internal server error.

The SPIDER section of the DMP notes that SPIDER data will be published through VESPA, but does not indicate who the data providers/publishers are or will be. Thus, it is not possible to verify that there is a data preservation plan for SPIDER data.

The statement in the DMP that "The SPIDER website and the repositories on the git service will be sustained after the end of the project" does not, in itself, constitute a plan. In particular, what is missing is a statement regarding who, if anyone, will provide maintenance or curation after the project ends.

### 4.2.3 Specific actions suggested

It would likely be helpful to new space weather users if one more update was done to the text on the PSWS site to reflect the final state of that project and to update any broken links persisting. It might also be helpful to new users to raise the visibility of the PSWS site from the SPIDER page, and vice versa (simply increasing the font size of the "Discover the PSWS website" text, for example, would help).

The DMP should be updated to clearly indicate the data providers (or, alternatively, the specific sections in Annex C of the DMP that apply) in the SPIDER case. In addition, a plan for maintenance or

curation (e.g., simple preservation, perhaps by containerization, in their current state without maintenance or further development) of the tools and services should be added to the DMP.

## 4.3 GMAP

The GMAP project enjoyed some surprise success with virtual training and conferencing, in that many more people were able to participate than would have been able to with strictly in-person workshops and meetings. Deliverables for the period included two significant documents: a basic training manual for planetary mapping and a detailed description of the Data Integration Portal.

### 4.3.1 Highlights

The forced shift to virtual training had an unexpected benefit, in that it created an opportunity to accommodate more training participants than would have been possible at an in-person session. For the presenters, attendance was not limited by the physical space available, and for potential participants the travel requirement barrier was removed. The shift to virtual training and collaboration has served the GMAP project well in mitigating the impacts of the pandemic travel restrictions. In addition to the training sessions, the “Basic Geological Maps” training document was published.

In addition to taking actions to strengthen ties between existing partners in the EU and China, new collaborations have been established with the US Geological Service, and with the Italian Geological Society and the Geological Survey of Italy through an agreement with the *Geological Field Trips and Maps* journal.

If a user can find and query the data access portal (see following), the results page is well laid out, and the landing pages for individual products are nicely detailed. The inclusion of citation formatting tools in those pages is an important precursor to integration of DOIs into the system.

The “Mapping Training Material: Basic geological maps” document was delivered as expected. It appears to be comprehensive and should prove very useful to the growing community.

### 4.3.2 Specific issues

While the GMAP portal site looks good, there are a number of issues with content:

- There appear to be no links to either data or to the public wiki site (mentioned in the GMAP report) anywhere on the portal.
- The “Documentation” page on the portal does not list any documents at all. This tab would be the place where new users would be expecting to find, for example, the basic training document.
- The first “News” banner is dated 24 November 2020, which is more than 18 months ago. The latest date on the news carousel is December 2021, which is more than 6 months ago.
- The “Tools” page offers only a link to the GMAP github site, which as of this writing contains 50 different repos. A potential user is given no indication of what sort of tools are available

by the website, and must resort to browsing through pages of one-line descriptions of repos hoping to find a description that might lead to something useful.

- The 2021 workshop videos are the only “Resource” listed, and they require a registration to view.
- Given the age of the “news” on the site and the high-level references to “Call #1”, it is not clear that a “Call #2” ever has or ever will happen.

The “GMAP Data Integration Portal” document (listed as a “demo” deliverable) talks about the “data integration portal” as though it exists. This does not appear to be the case. The listed entry point returns “Service not available”, as do the links to maps data access, the geoserver backend, and the Jupyter hub. The data link to the GMAP repository does work, in that it returns a search box, but provides no guidance for users who might be seeking data describing what holdings might be available or how to formulate a successful query. However, simple queries on target bodies do return relevant results, so the search interface appears to be intuitive to a reasonable extent. As noted in [4.3.1 Highlights](#), the content and layout of the result set are excellent.

While the annual report does rightly point to the developing collaborations as an element of sustainability, such dissemination does not itself guarantee preservation of the data. The GMAP section of the DMP notes that the basemap and input data files are already archival products curated by national archives and thus guaranteed a long life, but does not describe any actions to archive the derived products to a similar degree. If the derived products should not be considered “archivable” - that is, they are not of lasting significance or do not provide a foundation for further analysis over generational time spans - then it is reasonable to argue that the dissemination to collaborators provides sufficient redundancy to serve contemporary needs for what are, ultimately, ephemeral data products. Otherwise, a plan for long-term preservation of GMAP products should be added to the DMP.

### 4.3.3 Specific actions suggested

If the Data Integration Portal is supposed to be operational as described in the related document, then it needs immediate attention. If it is not, a prominent statement should be added to the document indicating when the various elements are planned for implementation. Allowing users to discover that elements are not operational by following bad links is the least desirable outcome.

The GMAP Portal should be updated to address the issues raised in the previous section. It is currently a dead end, leading to neither data nor documentation. In its current state, the GMAP Portal is a missed dissemination opportunity to reach users who might discover GMAP through the Europlanet site or search engines. At worst, it might be giving Europlanet visitors a very mistaken impression that the GMAP project is unproductive and largely inactive.

The “Mapping Training Material: Basic geological maps” document would benefit from technical editing to clean up some formatting inconsistencies, enhance readability of lists within sections (as, for example, in section 1.a *Approaches to Geoscientific Cartography*), and make sure captions and figures are properly aligned. Also, the edition of a glossary of mapping terms (e.g., equifinality, morphostratigraphic, chronostratigraphic, etc.) would be a very welcome addition for those new to



planetary mapping. This manual also repeatedly uses the term “shapefile” to indicate a file that represents a polygonal feature map, but does not indicate the format of the associated file. There is a “shapefile format”, developed by ArcView GIS and specifically designed to hold geospatial vector data, with well-defined, structured content and related file organisation. If the term “shapefile” is being used to refer to files in ArcView GIS shapefile format, then that should be made clear, even if this would be equivalent to a format recommendation. If it is not, then it would be better to use a generic description such as “vector format” or “polygonal format” rather than “shapefile”.

## 4.4 ML

The ML project had significant success in the past year with dissemination activities, and has made significant improvements to its portal site. Plans to integrate the first data products into VESPA and the first ML pipelines into SPIDER are on track to be completed in Year 3. Even though there is no VA-specific work package defined for ML, the ML project activities are achieving the overall Virtual Access goals.

### 4.4.1 Highlights

ML presentations were made at AGU and as part of the Open Planetary Lunch series; the first of four workshops with the fireball tracking community was held (virtually); and two ML pipelines were presented at EPSC 2021 - both of which are now publicly available on GitHub along with their accompanying tutorial or workshop repos. The ML team also organised sessions at both EGU and EPSC.

The beginning of integration of ML-produced data sets into VESPA is a significant positive step; as is the exploration of additional platforms for deployment (EXPLORE and ESA DataLabs).

### 4.4.2 Specific issues

The portal site layout is not very congenial to novice users. While it is easy for a novice to click through the content, the “General Introduction” and science case descriptions make liberal use of ML terms unlikely to be familiar to a novice. The introductory material that might demystify this terminology is only available by clicking a link on a page identified by the catch-all title of “Other tutorial and links”, and then the guides that might provide the terminology are links to books, not glossaries. Taken altogether, this creates an impression that the site is intended only for the already-initiated.

A number of the Jupyter notebooks provided as tutorials are too large to be displayed directly from the GitHub repo (the LMSU-Mercury\_boundaries demonstrator and IWF tutorial, for example). When downloaded and installed, the comments and tutorial remarks included in the notebooks assumed a familiarity with machine learning terminology and techniques that a novice user would not likely have, limiting their utility as tutorials. This was also true of the text in [https://epn-ml.github.io/tutorial01\\_supclass/intro.html](https://epn-ml.github.io/tutorial01_supclass/intro.html) site.

The repos in the GitHub site do not appear to have any consistent conventions for directory structure, naming, or documentation. Perhaps this is due to the nature of machine learning projects

at this point in time, but this seems like an excellent opportunity to lead the community in developing best practices, if not actual standards, for presenting the results of the science cases.

The DMP notes that “the data products, including the original data sets as well as trained models (where applicable)” will be published through VESPA. The intention of the “(where applicable)” comment is not clear, however. It seems to indicate that there would be cases where the data would be published to VESPA, but the trained models would not. That seems odd. In addition, the current VESPA “Dataproduct Type” list does not include anything corresponding to an ML model. No relevant file content or format standards are defined by VESPA. It is not clear how a user could discover ML models once they are published.

#### 4.4.3 Specific actions suggested

When updating the website, consider reordering or re-titling pages and sections to be a little more welcoming to novice users. For example, “Tutorials” is an excellent high-level title to see on the portal directory, but all the introductory information a novice user would need is collected under the “Other” banner at the bottom of the list. It might be better to divide the “Tutorials” section into two subsections: General tutorials for new users, and science case tutorials for those seeking to explore specific solutions (not necessarily in that order). Also, a glossary page of ML terms used on the site would be a welcome addition to the information for novice users.

Tutorial information on preprocessing has yet to be added to the current documents. While the document is noted as being “a work in progress”, there is no indication of what is intended to be added to the current tutorial vs. what is planned for inclusion in other deliverables. Preprocessing is, of course, a critical part of the application of ML to a problem. To emphasise the importance of this step to new users, it might be useful to have a visible placeholder for that information in either the current document or the list of tutorials for novices - wherever the information will eventually be available.

Consider breaking the longer Jupyter notebook tutorials into smaller modules that can be directly viewed through a browser rather than requiring downloading. This would not only encourage new users to try them out, but also helps to ensure that installation problems with the notebooks, unrelated to the machine learning pipeline itself, do not deter potential users.

In coming years it would be helpful in assessing impact and dissemination to see metrics related to activity in the repos as well as publication information and access numbers for VESPA activity related to the published data. The scope for ML dissemination and impact is potentially very broad. Any additional information that can be provided on aspects of dissemination unique to ML, for example, the deployment of SDAs to ESA Datalabs and similar platforms, would also be much appreciated.

The year 2 report includes descriptions for the science cases that discuss the science in detail but provide only superficial information about the ML process. These descriptions are nearly identical to the descriptions on the portal. Additional detail on the ML process, with figures, where appropriate, would be appreciated in both cases. Understandably, this may be something that can only be provided once the particular case has been fully developed.

With respect to standardisation of ML packages for dissemination and publication through VESPA, some response from the project is in order, even if that response is only an explanation of why it is impossible to take any steps to standardise the documentation, packaging, and deployment of the pipelines being produced. Such a statement should be made formally, in the DMP, so that the expectations of potential users of ML pipelines are properly calibrated. In particular, the question of discoverability in VESPA should be addressed.

## 4.5 Data Management Plan

The first version of the DMP was provided as a deliverable for this period. Some specific remarks are included in preceding sections where DMP issues overlap with other topics, like web site content or dissemination. This section collects the comments relevant to the DMP and its content.

### 4.5.1 Highlights

The VESPA part of the DMP is exceptional: Very clear, with well identified data types and detailed lists of associated responsibilities for each one of the VESPA contributors.

The acronym list was much appreciated, as were the references to the FAIR data principles.

### 4.5.2 Specific issues

These will be listed by project. Any issues that might affect DMP content noted in the preceding project sections will be repeated here in brief, for better reference when editing.

#### VESPA

None noted.

#### SPIDER

SPIDER data is published through VESPA, but the curators of the SPIDER data should be identified to ensure that the VESPA plan includes the associated responsibilities for those curators.

#### GMAP

There are several points that are unclear:

- Section 4.4.3, "Data types, formats, and standards" (bottom of page 15 in the PDF file) does not, in fact, describe any data types, formats, or standards. This very short section also ends in the middle of a sentence. It appears some significant text has been lost. This may be contributing to some of the confusion later in the section.
- In section 4.4.4.1, the term "raw data" is used to refer to data coming from the PSA and PDS archives; but then in section 4.4.5.1 the term "raw data" is apparently being used to refer to data that has been processed by GMAP, and the term "archived data" is apparently being used to refer to data from PSA/PDS. This leads to uncertainty in what is considered "raw data" whenever the term is used, and what organisation is responsible for preserving the "raw data".

- Section 4.4.9 refers to “data and maps published on the GMAP data archive (ESA Guest Storage facility, INAF, UNIPD or other institutional data repositories)”. These institutions (along with VESPA) are also mentioned in an earlier section as supporting discoverability. But there is no indication of preservation plans for these organisations.
- Although the word “exemplary” in the title of 4.4.15 presumably indicates that the description in that section should serve as a good example, the implications of the word “exemplary” and the way the case is presented strongly imply that all data will go to ESA GSF, which is not the case given the information in section 4.4.9.
- There is no estimation of the total number of maps and related data files likely to be produced as part of the GMAP project, nor in how they are to be distributed among the named repositories (will all repositories have copies of all maps, for example).

## ML

While the descriptions of the input and processed formats are clear, there is no mention of the formats for the trained models themselves. Without standardisation for this in the field, some discussion of the anticipated formats and how they might affect usability in the short and longer term would be appropriate. In particular, while the data might be interoperable, it is not at all clear that the models themselves will be.

Section 4.5.2 states, “original data sets as well as trained models (where applicable), are aimed to be published in VESPA”, but the significance of the “where applicable” comment is not clear. This is noted in Section 4.4.2 Specific issues, above.

That same section concludes, “The ML Portal and the repositories on the git service will be sustained after the end of the project.” However, there is no indication given of how they will be sustained, and whether that includes maintenance and support, or merely access to orphaned code and documents.

### 4.5.3 Specific actions suggested

An additional round of editing to clarify the points raised in the preceding section is suggested.

The approach taken by VESPA to list the responsibilities of individual institutions acting as repositories is a reasonable one to take where these specifics are missing for other projects.

## 5. Any Additional comments and suggestions

The review board noted that a number of the portals and related services are not being served via secure HTTP. Trusted SSL certificates can be obtained at no charge and renewed automatically. They provide a minimum level of assurance for users that the website is what it claims to be, especially for users working over the potentially hazardous environment of public wifi. Implementing at least that minimum level of security will also prevent new users from having to run the gauntlet of warning messages and roadblocks generated by browsers for websites that do not run secure HTTP services.

The board strongly recommends that all portals and services be moved to secure HTTP service as soon as possible.

The patience on the part of Europlanet Society and the VA projects in waiting for this report, repeatedly delayed as the reviewers dealt with ongoing pressures of last-minute schedule changes and the severe schedule disruptions caused by virtual meetings in distant time zones, is greatly appreciated. We all hope for more predictability in the coming year.