EUROPLANET 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:	28 Feb 2022
Nature[1]:	R
Dissemination level[2]:	PU
Work package:	WP01
Lead beneficiary:	Constructor University
Contributing beneficiaries:	External review
Document status:	Final

Start date of project:	01 February 2020
Project Duration:	54 months
Co-ordinator:	Prof Nigel Mason, University of Kent

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. Reference period

The reference period for this report is 12 months, from February 2022 through January 2023.

2. Composition of the review board

Anne Raugh, University of Maryland
Kiri Wagstaff, AAAS Congressional Fellow in Artificial Intelligence
Benjamin Lynch, University of California - Berkeley
James A. Skinner, Jr. United States Geological Survey

3. General Comments

1. 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.

Yes. Each VA project has met its deadlines as revised for the disruptions caused by the 2020 pandemic. Deliverables were provided in good order; workshops, where applicable, have been resumed; and collaborations have continued.

2. 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.

The VA projects have all had success in making an impact on their communities through the appropriate combinations of active participation in meetings and conferences, citations in the research literature, involvement of existing and new users in workshops, and international collaborations.

3. 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.

Yes. The projects have met their goals for dissemination as appropriate to each project. All are producing documentation for users; all have been present at conferences where their user communities are represented. Training materials developed in the initial years have been refined and/or expanded to better serve users. New services have been brought online and existing services have been improved. In particular, collaborations between both individuals and agencies have grown in number and strength.

4. Specific per-VA Comments

4. 4.1 VESPA

VESPA continues to expand its services, support progress in standards with updated services, and provide leadership in best practices for data services through their approaches to, for example, data management plans and application of DOIs. Work continues in the international community creating and promoting the adoption of standards and vocabularies as a primary focus of sustainability activities.

4.1.1 Highlights

More than 60 data services are deployed, with updates and upgrades to older services made by multiple participants. In particular, upgrades to the Spectro_asteroids and SSHADE spectral archives have substantially enlarged those complementary services; and the Mars Global Client Model service and documentation are now online. A number of existing services were upgraded to comply with the new EPN-TAP standard.

Regular workshops have resumed and have contributed to the creation of new services.

The VESPA portal is seeing steady use, with peaks noted when new services are brought online.

4.1.2 Specific issues

While the GitHub tutorials collection seems very helpful, it appears to be languishing. No issues have been resolved; incomplete tutorials still contain placeholder text; etc. These do not seem to overlap with the tutorials presented on the support portal page although the portal text implies they might.

4.1.3 Specific actions suggested

A review of the Support Site links and wiki documentation would be in order, to ensure that user support information is as complete and current as possible as end-of-funding approaches. For example:

- The "AOs and Workshops" wiki page is about 4 years out of date (refers to 2020RI)
- "Setting up a VESPA Service in EPN2024" (VESPA- WP6-3-011-TD-v3.0(32)), on "Data Services" web page, is marked as a "work in progress" but hasn't changed since 2021.
- On http://www.europlanet-vespa.eu/develop.shtml, the link to the PDS4 EPNTAP dictionary leads to a 404 page.
- etc.

5. 4.2 SPIDER

SPIDER is on schedule and is in a good position to complete its remaining milestones on time.

4.2.1 Highlights

The exospheric run-on-demand services are up and running.

The SPIS plug-in tutorial is finished and includes a very complete set of screenshots to support new users.

All planned services are now operational at some level.

4.2.2 Specific issues

It seems misleading to continue to refer to a run-on-demand service for "some Galilean moons" when in fact only Europa is included in that service. Yes, one is technically "some", but Europa is a significant target and it is worth mentioning it by name for discoverability. The reference to "some" might also raise expectations in new users that lead to immediate disappointment on discovering that "some" = "one".

The C7 task, “Planetary magnetospheric particle tracing runs-on-requests”, is not a run-on-request service, at least not in the same sense as the C5 Exospheric Model service. Again, it seems misleading to call them both “run on request” when the C7 service requires downloading, unpacking, and configuring software that also requires a separate Matlab licence to execute. The package does seem to be complete, in that there is documentation to assist in setting up and running the code.

The E2 “Hermean and Jovian magnetospheric simulation database” is reported as partially complete (only Mercury is included), however, the SPIDER website as of this writing indicates that this entire task is still “In development, operational end of 2022”. It appears that the Mercury data is in AMDA, as reported, but there is no mention of this on the SPIDER “Database” page. Neither is it mentioned in the “Announcements” listed on the AMDA home page.

4.2.3 Specific actions suggested

The SPIDER website needs some minor updating to address some inconsistencies and omissions:

- The description of the C5 task should refer to “the Galilean moon Europa” rather than “some Galilean moons”, as this is the only moon exosphere modelled. Alternatively, simply add the name “Europa” to the description, so that search engines will find it.
- The C7 title should not use the same “run-on-request” terminology for a software installation package as the C5 title does for an interactive service.
- Consider adding a link to the C7 section of the web page to a brief document summarising what the “here” link actually links to.
- The current state of the E2 task, and the location of the data that has been delivered, should be indicated on the “Databases” page.
- The “Presentations & News” page has not been updated since 2020 and two of the three presentations are not viewable because of their privacy settings. It is generally better to have no “news” page than a “news” page that is clearly not being updated. A project that has no “news” to report in over three years is generally assumed to be either dead or dying.

6. 4.3 GMAP

GMAP documents, products, and workshops continue to make significant contributions to the mapping community. As the Research Initiative enters its final year, GMAP has developed communities and promoted relationships that will support continued cooperation and use of the materials and datasets created.

4.3.1 Highlights

The Winter School mapping workshop has become a recognized community resource that includes a diversity of presentation and participation.

The GMAP web portal is a well-maintained resource for geologic map information, and the associated wiki successfully provides various sets of documentation - many linked for searchability.

GMAP has developed synergy with other EU projects for providing base mapping services, as well as cooperation with the USGS to plan common actions for conferences, workshops, meetings, and publications.

4.3.2 Specific issues

None noted.

4.3.3 Specific actions suggested

None noted.

7. 4.4 ML

The ML team's primary goal during this period was to produce a tutorial on machine learning with some basic how-to advice. They have made very good progress on what has grown to be a book. Some parts are not yet written (Introduction, Chapter 3 on Supervised Regression, Chapter 7, etc.). As noted by the authors, there are other existing books and references that can provide the basic machine learning background needed to appreciate and make use of the book's main contribution, which is a collection of use cases in which machine learning is used to analyse scientific data.

4.4.1 Highlights

The initial release of the *Europlanet Machine Learning Book* was published. This book presents ten use cases for machine learning, primarily employing planetary science data. There is great potential for others to experiment with and emulate the cases presented in this resource.

In addition to the book itself, the code for the published use cases is provided via GitHub and Jupyter notebooks. These repos will be helpful for those learning from examples.

(We note that the content of the *Europlanet Machine Learning Book* is not yet complete: Four use cases are only partially documented - without methods or results - and some content is still in bullet point format at the end of Chapter 6. We look forward to seeing these complete in the final release.)

4.4.2 Specific issues

The book appears to be a loosely organised compilation of output of multiple Jupyter notebooks. While the Jupyter notebooks are great as a learning tool, they do not automatically produce a nice, readable book. Currently the book is almost 350 pages long despite being only about 50% complete in this initial release. While there is a lot of good content, overall the book is too long and, in its static PDF form, suffers from a fatiguing quantity of detail in both description and output listings. For example

- There are many pages of output (like epoch timestamps) included verbatim.
- There are >130 pages of plots in Chapter 6 without discussion or commentary
- Sections 11.6 and 14.2 (and others) include tracebacks from interrupted or failed executions
- Section 17.2 includes a long printout of progress bars

Some of the plots do not render well in the PDF book, especially of the magnetic boundary crossings (sec. 6.5.5)

4.4.3 Specific actions suggested

The book would benefit from a rigorous technical edit:

- Go through the book and select only the content that is informative to the reader. Exclude code dumps and error messages (unless they are illustrative of advice you are providing). Select a representative plot instead of pages of the same kind of plot. All content should be accompanied by text discussing important points (and ML tips) for the reader. The resulting book will be much shorter and have a higher impact. It can include pointers to the full Jupyter notebooks for those who want to work through the details.
- Check all plots for legibility.
- Add attribution for the author of each use case's code.

5. Any Additional comments and suggestions

It is understood that the ability of any project to act on specific suggestions provided above is significantly hampered by the realities of the linearity of time. As each project prepares for end-of-funding, the focus is shifting to sustainability and production of the materials and collaborations that will ensure that the communities will continue to benefit from the work done and progress made. This is appropriate, and the focus in each case seems to be well-placed.

One technical note: The lack of HTTPS implementation across all services is an ongoing problem. Browsers are aggressively blocking communication with unsecured HTTP services, and the insecure protocol is increasingly being identified as a red flag signalling likely sources of malware and similarly malicious activity. Continuing to cling to the HTTP protocol does not reflect well on the service being offered.