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 D8.2

Deliverable Title:	GMAP Data Integration Portal
Due date of deliverable:	May 31, 2021
Nature[1]:	R
Dissemination level[2]:	PU
Work package:	8
Lead beneficiary:	UNIPD
Contributing beneficiaries:	JACOBSUNI, DLR, UNCH, CBK-PAN, WWU
Document status:	Final

Start date of project:	01 February 2020
Project Duration:	48 months
Co-ordinator:	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)

Executive Summary / Abstract:

The GMAP Data Integration portal is described with its components, data access services, documentation and tools to support virtual access (VA) and perform geological mapping based on established GMAP standards.

The GMAP portal provides an entry point for both data access and data submission, as well as separate support documentation and tool access directory for geologic mapping projects to start and progress. The data submission is primarily aimed at those creating data (geological mappers), while the other tools and facilities are community-facing services available to both geological mappers and the more general planetary geology and mapping communities. The tools have been provided either directly from the GMAP community or collected from externally available sources. Documentation for use of the tools and services has been compiled from both internal and external sources. Such documents and sources will also be of interest for the general planetary science community, e.g. for image processing, basemap preparations for outreach or amateur planetary scientists.

Table of Contents

1. Acronym list

Acronym	*Description*
CRS	*Coordinate Reference System*
ESA	*European Space Agency*
FGDC	*Federal Geographic Data Committee*
JRA	*Joint Research Activity*
MARSIS	*Mars Advanced Radar for Subsurface and Ionosphere Sounding (ESA MEX experiment)*
MEX	*Mars Express (ESA mission)*
MRO	*Mars Reconnaissance Orbiter (NASA mission)*
NASA	*National Aeronautics and Space Administration*
OGC	*Open Geospatial Consortium*
PDS	*Planetary Data System*
PLANMAP	*PLANetary MAPping (H2020 Space project - https://planmap.eu/)*
PSA	*Planetary Science Archive (ESA archive)*
REST	*REpresentational State Transfer*
SHARAD	*SHAllow RADar (NASA MRO experiment)*
USGS	*United States Geological Survey*
VA	*Virtual Access*
WCS	*Web Coverage Service (standard, OGC web service)*
WFS	*Web File Service (standard, OGC web service)*
WMS	*Web Map Service (standard, OGC web service)*

2. Introduction

The modern planetary geologic mapping process is based on established principles and digital mapping practice and tools (Nass et al., 2020). In addition to basemaps,which are the focus of a future JRA deliverable (D9.3), the process requires standards (Nass et al., 2020), documentation, guidelines (e.g., Tanaka et al., 2010; Asch and Vinnermann, 2011), templates and methods of data sharing and publishing Spatial Data Infrastructures (Skinner et al., 2019; Radebaugh et al., 2020; Laura et al., 2017), for both scientific and technical review and interim or final publishing of products (Nass and van Gasselt, 2013).

The GMAP portal provides an entry point for both data access and data submission, as well as separate support documentation and tool access directory for geologic mapping projects to start and progress, following the basic process highlighted in the 2021 GMAP/PLANMAP Winter School[1], and building up on the metadata standards of PLANMAP, for both map-wide metadata and vector metadata, i.e. mapping attribute table. The data integration portal is thus supported by a set of progressively updated documentation and tools (see Table 1).

3. Data integration portal components

The data integration portal has its main entry point on the map browser and data submission system[2]. The various services are included in Table 1 and briefly described in the following subsections.

Table 1: Components of the Data integration portal and associated documentation entry points

Component	***Access URL***	***Type***
Data Integration Portal entry point	*https://cloud.europlanet-gmap.eu*	*Service*
File data access	*https://data.europlanet-gmap.eu*	*Service*
Map data access	*https://maps.europlanet-gmap.eu*	*Service*
Geoserver backend	*https://geoserver.europlanet-gmap.eu/geoserver*	*Service*
Jupyter Hub	*https://jupyter.europlanet-gmap.eu*	*Service*
Templates	*https://github.com/europlanet-gmap/gmap_metadata*	*Templates*
Documentation entry wiki page	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/*	*Documentation*

[1] https://www.planetarymapping.eu/
[2] https://cloud.europlanet-gmap.eu/

2. Services

Data integration portal entry point

The data integration portal provides the user interfaces and back-end mechanisms for the publication and access to data packages related to the project. It will also serve any products inherited from previous projects (i.e., PLANMAP).

The portal is composed of a set of services interconnected to build a workflow for data publication and management (Figure 1, 2). The primary goal of the services is to provide easy-to-use interfaces both for the data producer and the consumer. Their implementation design is guided by flexibility and extensibility.

Flexibility relates to accepting different formats of data and not-so-complete metadata around a dataset. This is important because high-level planetary data can be delivered in a variety of different formats (not only purely geographic information but also 3D models, spectral cubes and other derived products; e.g., see Rossi et al., 2021; Caravaca et al. 2020; Pozzobon and Penasa, 2021; Brandt et al., 2020).

Furthermore, it must be recognised that data providers (i.e., researchers and mappers in general) work and create habits around certain data formats and proprietary software, so that no single organization of the work can be expected. Flexibility is thus needed to create as little friction as possible between the working habits of the provider and the subsequent dataset delivery and hosting. Flexibility might also favour the testing of new practices which may better fit new datasets or science topics.

Metadata play an important role in filling the gap between the content of the provided dataset and its specific role (e.g., whether a file represents a geological map or another type of content). Services that are too restrictive to what (data/metadata) can be accepted may become an unnecessary obstacle for data to become public, which is the primary goal of the project. Hence, requiring only basic documentation in terms of overall authorship and description of the dataset (e.g., as used in Zenodo[3]) is preferable.

As data is collated (GMAP systems) and according to the state data/metadata provided, different services can be enabled for that dataset, to provide higher-level interfaces to access the data (besides pure search and download). That is where *extensibility* is important: the set of services around a data package can expand according to (1) data and metadata status, and (2) the actual software implemented on our premises. In view of this, services are designed to be loosely coupled so they can grow fairly independently. An additional role of the data integration portal is to provide an easy-to-use interface that will allow the user to improve the data and metadata to make the product integrable with the rest of the architecture.

Ease of use relates to how intuitive an interface is to the user, or how simple/complex it is to learn how to use it. This is naturally related to the previous

[3] https://zenodo.org/

experience the users bring with them. We have chosen to provide simple, non-technical interfaces for data access such as a simple files-server (which every single individual is used to) and an interactive maps-viewer, as well as more technically bounded interfaces such as OGC/W*S services for programmatic query and on-demand data access.

Being able to provide different services for the same data collection in a loosely coupled architecture challenges the synchronisation of such data and metadata across the services. Data consistency across different views and access methods is a key aspect to consider, especially when the data providers are capable of updating their data packages with more data or better metadata (a capability provided). To achieve this goal, the software behind each service must be largely volatile and autonomous, meaning it should understand data packages and recognise modifications to the set on its own.

1. Data submission

Data packages are uploaded through the data submission page (Figure 3), which provides a simple interface with the list of data files associated. The data files are associated with the package by the user in the very same interface by uploading the files (see figure below).

Not all fields in the form are mandatory; only the very basic information necessary to give the package a (unique) GMAP ID must be inserted at this stage, namely the target body and the package name. Some fields are automatically extracted from the provided data, for instance, the enclosing bounding-box and the associated coordinate reference system (CRS). Non-mandatory fields and metadata that could not be automatically extracted from the data itself can be then added by the user to provide more detailed information over the package content.

Since we do want any reasonable (GMAP) data package to go public, but also encourage the creator to give the best of data/metadata content, the system is designed to implicitly reward the provider. The reward is done through (1) services associated with the packages (e.g., linked or not to maps-viewer), and (2) rank in the search results: higher scored packages are listed first (i.e., packages with proper inserted metadata, i.e., without missing important information). To avoid being judgmental over the quality of data/metadata provided, scoring a data package is applied over the percentage of metadata fields filled by the creator. Moreover, key (mandatory) fields like bounding-box or ancillary data products enable a package to be hyperlinked to higher-level services or external resources (e.g., USGS/PDS source data archives).

Figure 1: The GMAP integration portal entry page, a completely new development of GMAP, hosting exemplary PLANMAP H2020 products (see Table 1).

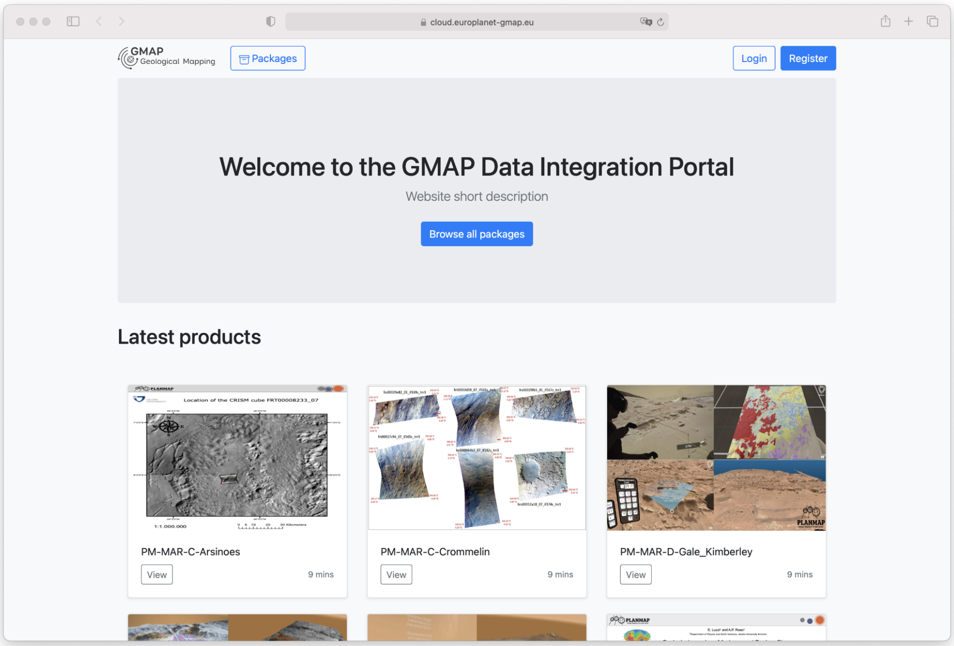


Figure 2: Exemplary map-wide metadata entry for one GMAP Portal entry, originally from PLANMAP data (see Table 1).

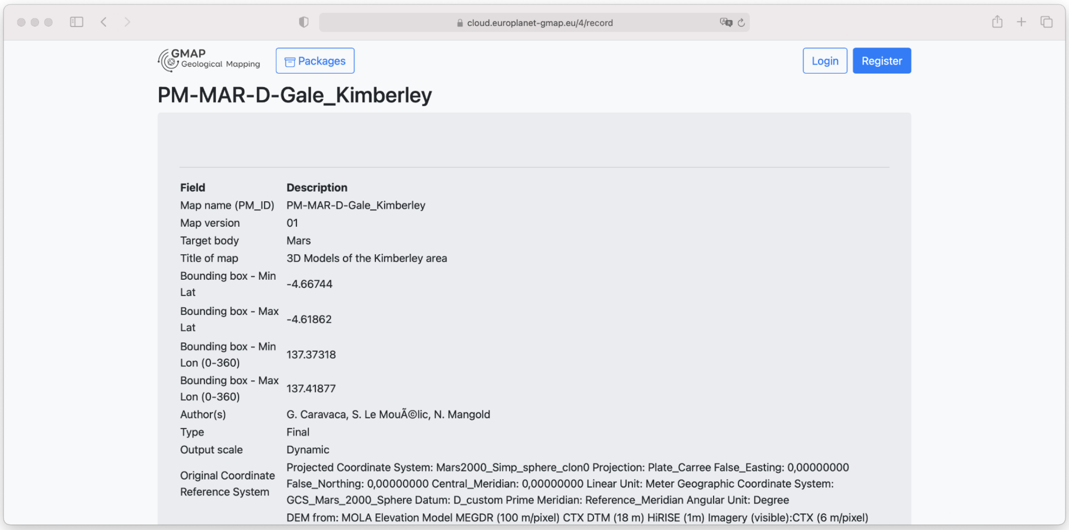
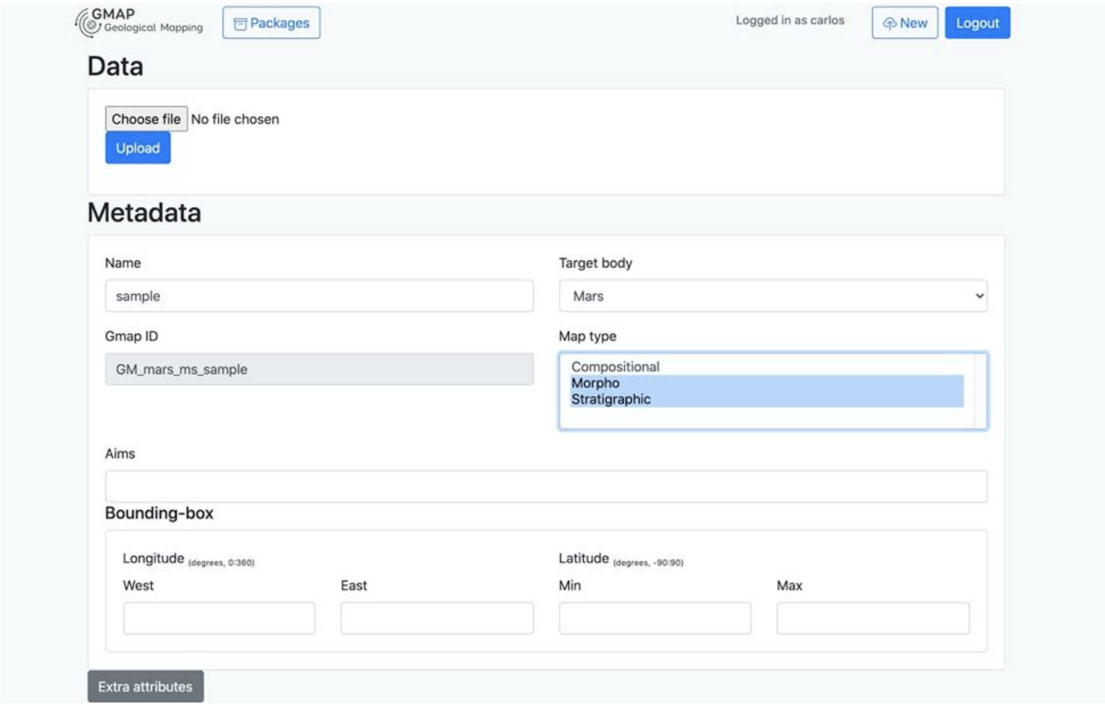


Figure 3: The GMAP integration portal data submission page (see Table 1).

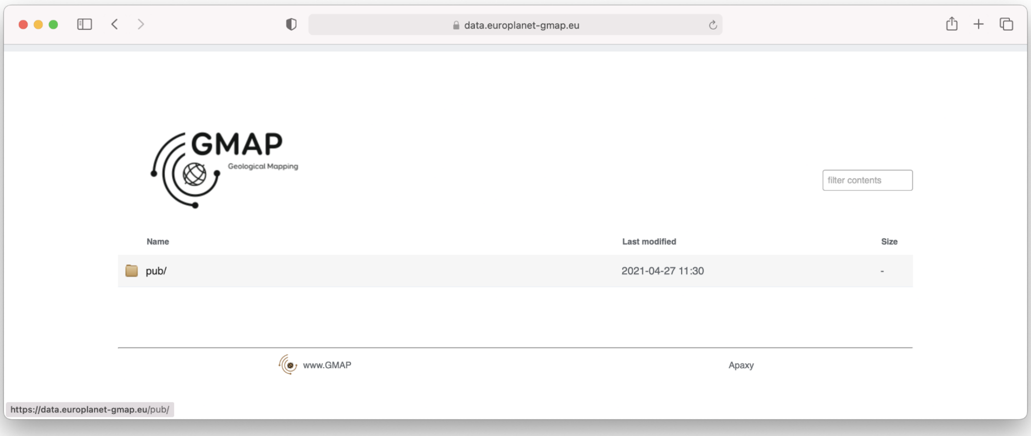


File data access

The simplest interface for data access is a file-server, where every individual knows how to navigate through a tree of directories and files (Figure 4). This is a fundamental interface, well known by the planetary community and often used to serve data from PDS nodes. A file-server is not only a common interface to interactively navigate through, it is also a straightforward way of batch downloading data (i.e., anybody without much knowledge of computing can download entire datasets with the help of a file downloader).

Data packages are organised after their respective target body, which is a well-defined non-overlapping property among the packages. The user then finds data packages organised in alphabetical order.

Figure 4: The GMAP file access interface, inherited and adapted from PLANMAP H2020 (see Table 1)

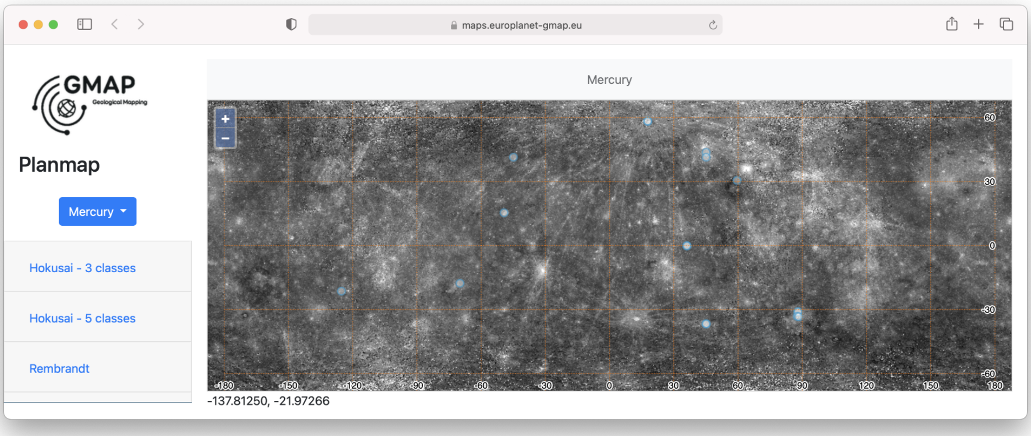


Map data access

An interactive view for data packages exploration presents the packages' location on top of the corresponding target-body base map (Figure 5). As previously discussed, an incremental approach regarding data/metadata status makes it possible to serve the products fulfilling the needed criteria (i.e., at least bounding-box information) by the maps-viewer, which is a service presenting data packages as an interactive map.

Through the maps-viewer, besides having the overview of data spatial distribution, the user is also informed about the data content. Maps-viewer is not a data analysis interface, but an interface for interactive data discovery. From here, the user is redirected to the data package content itself.

Figure 5: The GMAP map access interface, inherited and adapted from PLANMAP H2020 (see Table 1)

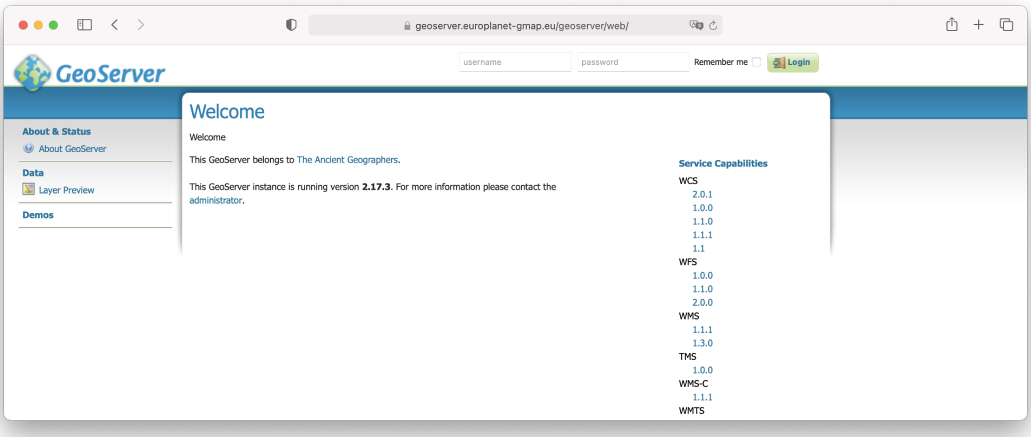


GeoServer backend

When users/data providers submit data content fulfilling spatial dataset attributes, this information is used to provide services that respond to spatial queries. The technical solutions used are commonly used and well known by the community (i.e., the standard OGC Web Services).

Spatial datasets are served by the GeoServer (Figure 6), which implements OGC's WMS, WFS and WCS services all well known and directly usable with most current GIS software. GeoServer has a number of advantages compared to other spatial databases: it is stable, well documented, portable, and provides graphical as well as command-line (i.e., RESTful) interfaces. All these features are not only important for the user experience, but also from a system-administration point of view.

Figure 6: The GMAP GeoServer backend to host and serve basemaps and mapping products (see Table 1)

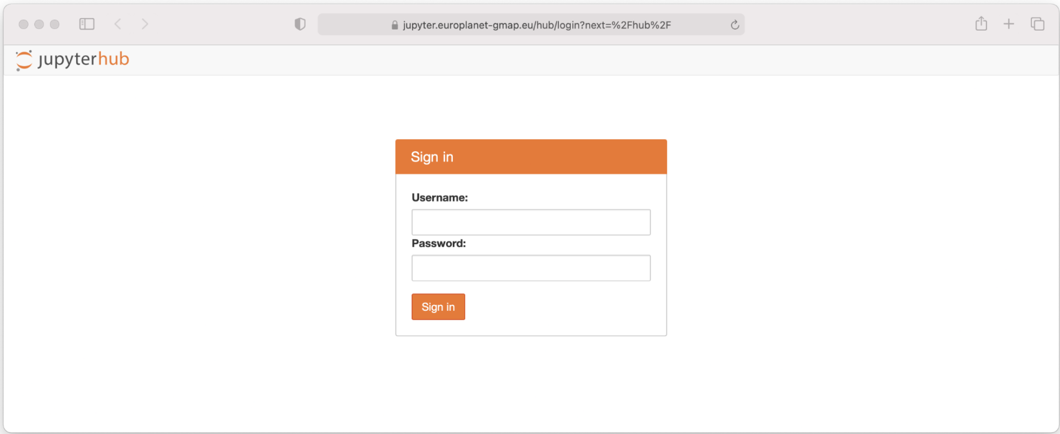


Jupyter Hub

Jupyter Hub is an environment for users to analyze data and interact with data in a lower-level, using Python tools and the libraries developed by the team. Although all tools (and libraries) are publicly available, and so every willing user can set up their own environment, we provide a preset environment (1) to simplify the workflow of users not familiar with computing technicalities, (2) to guarantee a minimal set of computing resources and stability for everyday tasks.

The Jupyter Hub environment authenticates users through our project GitLab/GitHub's registered set of users - it is not open to any public user for security and stability reasons. The environment provides all the necessary common libraries for geo-processing, data access tools and ancillary data provided by the project (Figure 7).

Figure 7: The entry page of GMAP Jupyter Hub, to host both data access and analysis functionalities (see Table 1).



3. Tools

Some of the tools available to the Virtual Access (VA) community, in addition to commonly used Open Source software for planetary data processing and analysis (see for a review Nass et al., 2020 and Appendices) have been developed in previous projects such as PLANMAP H2020, or are being developed specifically for GMAP. A list of external software-related documentation pages for relevant tools for performing actual geologic mapping is provided on the GMAP wiki (see Documentation section).

Mappy (PLANMAP heritage)

Mappy (Penasa et al., 2020) is a QGIS plugin developed in Python that simplifies the process of creating geometrically sound geological maps by providing easy-to-use processing tools that make it possible to simplify the mapping process. The tool also serves the objective of fostering the adoption of good practices, which can, in turn, provide better datasets and improve the mapping experience.

This is accomplished by introducing mappers to more advanced geoprocessing methodologies by using simplified tools. Optimal approaches for geological map production are indeed well known but are rarely employed due to the additional and unintuitive processing that they require. Mappy is designed to streamline all the complex operations needed for map production and is also associated with theoretical documentation so that the same approaches can be replicated by using any GIS environment of choice. A specific documentation page is provided for this QGIS tool (see Documentation section).

Symbology

The Symbology QGIS add-on (Frigeri, 2020) introduces QGIS symbols useful in geological mapping. Styling for points, lines, fill patterns and gradients are included. Currently the library includes symbols defined by the Federal Geographic Data Committee (FGDC), but the project is structured in such a way that other libraries from different mapping authorities/institutions can be easily added (see also Nass et al, 2010, 2020). A specific documentation page is provided for this QGIS tool (see Documentation section).

3D Radar sounder tools (MARSIS and SHARAD)

A suite of tools developed by GMAP is provided (Table 2). Those are instrumental to perform 3D geologic mapping and modelling using subsurface data derived from planetary sounding radars, either on Mars or the Moon, in perspective for any Solar System body covered by relevant radar sounder data using a similar approach (see also Penasa and Pozzobon, 2020; Penasa et al., 2020; Nodjoumi et al., 2021).

Table 2: Initially released GMAP documentation pages to support new VA users.

Tools	***Code access***
Mappy (PLANMAP)	*https://github.com/europlanet-gmap/mappy*
QGIS Symbology	*https://github.com/europlanet-gmap/geologic-symbols-QGIS* *forked from original https://github.com/afrigeri/geologic-symbols-QGIS*
Sharpy - SHARAD seg-y converter (PLANMAP)	*https://github.com/europlanet-gmap/sharpy* *forked from https://github.com/planmap-eu/sharpy*
MARSIS PDS downloader	*https://github.com/europlanet-gmap/MARSIS_EDR-RDR_PDS-Downloader*
MARSIS RDR/EDR reader and seg-y converter	*https://github.com/europlanet-gmap/MARSIS-xDR-READER*

4. Documentation

The documentation accompanying the GMAP Data Integration Portal (Table 3) is hosted primarily on the GMAP wiki[4] (Table 1) and sources are placed there or in

[4] https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/

related GitLab[5] or GitHub[6] repositories (see also Table 2), as needed. External resources, e.g., from USGS, NASA PDS, ESA PSA, will be maintained and linked from the GMAP wiki documentation system.

Table 3: Initially released GMAP documentation pages to support new VA users.

Page	**Access URL**
CRS Simple Guide	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/CRS%20simple%20guide/*
GMAP naming conventions	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/GMAP%20naming%20conventions/*
Map-wide metadata	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/Map-wide%20metadata/*
Vector mapping fields	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/Vector%20mapping%20fields/*
Mappy simple guide	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/Mappy%20simple%20guide/*
Symbology simple guide	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/Symbology%20simple%20guide/*
Layout guidelines	*https://wiki.europlanet-gmap.eu/bin/view/Main/Documentation/Layout%20guidelines/*
External resources	*Maintained (also with support of the external community) on https://github.com/europlanet-gmap/awesome-planetary-geology*

CRS Simple guide

The documentation page provides basic information on CRS basics for geologic mappers choosing a map projection and planning basemap preparations. References and links to external resources (such as USGS Astrogeology) are provided.

GMAP naming conventions

GMAP inherits naming conventions from PLANMAP, with a system of substrings composing what is the unique identifier of a GMAP product. A brief description of the naming process and specific aspects, as well as a very brief description of the map types identified so far, is included in the GMAP naming conventions page. References and links are also provided.

GMAP map-wide metadata

Map-wide metadata are provided for describing each GMAP product to come, similarly to those from the USGS map products[7]. Some of these metadata are basic geometric info, such as bounding box, some are more specific, including unit names,

[5] https://git.europlanet-gmap.eu/
[6] https://github.com/europlanet-gmap
[7] e.g. https://astrogeology.usgs.gov/maps/

scientific references, for example to published papers related to the map, etc. The initial version of GMAP map-wide metadata is entirely inherited from PLANMAP. It might be updated in the course of the project with additional items. Map-wide metadata can also be used for data discovery.

GMAP Vector mapping fields

A set of predefined attribute table fields, inherited from PLANMAP, is provided and described. The vector mapping fields and the related geopackage templates (i.e., empty vector files for a starting mapping project) are intended to be used together with Mappy, although the agnostic attribute table can be used with any other mapping approach, regardless the vector topology and process adopted.

Mappy simple guide

The recently released Mappy tools are going to be heavily used within GMAP due to the underlying robustness, ease of use and flexibility for performing basic and advanced digital geologic mapping tasks. The page includes a subset of Mappy's documentation and points to relevant GitHub repositories and related resources, see also Penasa et al. (2020).

Symbology simple guide

The Symbology library, provided as an add-on to QGIS, includes symbols and patterns useful in the Earth and planetary geological mapping. Styling for points, lines, fill patterns and gradients are included. Currently, the library includes more than 100 user-contributed symbols and patterns defined by the FGDC for planetary geologic mapping, but others can be added in the course of GMAP activities.

GMAP Map layout guidelines

Guidelines and - later - QGIS layout templates will be provided. While VA mapping users can choose to use any Open Source or proprietary GIS system, GMAP supports the use of Open Source and accessible QGIS-based solution, thus, also the future map layout templates will follow this path.

4. Plans and future iterations of the GMAP data integration portal

The Portal and related services and documentation will be continuously updated and populated with content from VA users, external projects contributing to GMAP with maps or basemap services, partly addressed by the future JRA D9.3 deliverable and subsequent ones.

5. References

Asch, K. & Vinnemann, C. (2011): Guide for compilation of geoscientific maps. BGR (Hannover). 50 p

Brandt, C. H., et al. (2020) Planmap data packaging: lessons learned towards FAIR planetary geologic maps, EGU2020-18839 https://doi.org/10.5194/egusphere-egu2020-18839 , EGU General Assembly 2020,

Caravaca. G., and the PLANMAP Consortium (2020) D5.2, 3D products (geomodels) of the merged GIS and maps of Gale crater, PLANMAP deliverable, available online at https://wiki.planmap.eu/display/public/Deliverables

Frigeri, A. (2020); gsymblib: Geologic symbols library and development for QGIS. EGU 2020, Vienna - held online. DOI:10.5194/egusphere-egu2020-22625.

Laura, J., et al (2017) Towards a Planetary Spatial Data Infrastructure, ISPRS International Journal of Geo-Information, 6(6), 181, doi:10.3390/ijgi606018.

Nass, A., van Gasselt, S. (2013) Archiving and Public Dissemination of Planetary Geologic and Geomorphologic Maps. In: Proceedings, Seiten 17-24. Sharing Knowledge Symposium, 23. Aug. 2013, Dresden, Deutschland

Nass, et al., (2020) Standard definition Document 1st iteration, available online at https://www.europlanet-gmap.eu/about-gmap/deliverables .

Nodjoumi, G., Guallini, L., Orosei, R., Penasa, L., and Rossi, A. P. (2021) New open source tools for MARSIS: providing access to SEG-Y data format for 3D analysis., EGU General Assembly 2021, online, EGU21-4031, https://doi.org/10.5194/egusphere-egu21-4031 , 2021.

Penasa, L., Pozzobon, R. (2020) D6.2, 3D geo-models based on multiple datasets of the Moon (implicit or explicit modelling), PLANMAP deliverable, available online at https://wiki.planmap.eu/display/public/Deliverables.

Penasa, L., Frigeri, A., Pozzobon, R., Brandt, C. H., De Toffoli, B., Naß, A., Rossi, A. P., and Massironi, M.: Constructing and deconstructing geological maps: a QGIS plugin for creating topologically consistent geological cartography, Europlanet Science Congress 2020, online, 21 September–9 Oct 2020, EPSC2020-1057, https://doi.org/10.5194/epsc2020-1057 , 2020

Pozzobon, R., Penasa, L. (2021) D6.3, 3D geo-models based on multiple advanced datasets of Mercury (explicit modelling), PLANMAP deliverable, available online at https://wiki.planmap.eu/display/public/Deliverables.

Radebaugh, J., et al. (2020) Maximizing the Value of Solar System Data through Planetary Spatial Data Infrastructures, white paper submitted to the 2023–2032 Planetary Science and Astrobiology Decadal Survey, https://arxiv.org/pdf/2008.06171.

Rossi, A. P. Brandt. C., and the PLANMAP Consortium (2021) D7.7, Final release of data/code/products as VESPA compliant data, PLANMAP deliverable, available online at https://wiki.planmap.eu/display/public/Deliverables.

Skinner, J.A. Jr., et al. (2019) Planetary geologic mapping—program status and future needs: U.S. Geological Survey Open-File Report 2019–1012, 40 p., https://doi.org/10.3133/ofr2019101

Tanaka et al (2010) Planetary Geologic Mapping Handbook. USGS, 24 p.