euroPLANET2024
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 D8.8

Deliverable Title:	Geologic mapping integration report and published data
Due date of deliverable:	30th April 2022
Nature[1]:	Report
Dissemination level[2]:	Public
Work package:	WP8
Lead beneficiary:	UNIPD
Contributing beneficiaries:	JacobsUni, WWU, CBK-PAN, UDA
Document status:	Final

Start date of project:	01 February 2020
Project Duration:	48 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)

Executive Summary / Abstract:

The status of the GMAP Data Integration Portal is provided, with current capabilities and data hosted or linked, as well as future developments and data publication plans. The portal uses Open Source standards and tools, based on industry-standard best practice. Interoperability with similar systems on the United States Geological Survey (USGS) side are tested and they are going to be further developed and deployed throughout the course of the Research Infrastructure.

Table of Contents

Contents

List of acronyms and abbreviations

Table 1: Acronyms and abbreviations

Acronym	Description
API	Application Programming Interface
CERN	European Council for Nuclear Research
CUGB	China University of Geosciences Beijing
ESA	European Space Agency
FAIR	Findable, Accessible, Interoperable, Reproducible

GMAP	Geologic MAPping of Planetary Bodies
GUI	Graphical User Interface
JRA	Joint Research Activity
NA	Networking Activity
MOST	Ministry Of Science and Technology
NASA	National Aeronautics and Space Administration
PDS	Planetary Data System
PLANMAP	PLANetary MAPping Project
PSA	Planetary Science Archive
RDM	Research Data Management
USGS	United States Geological Survey
VA	Virtual Access

Introduction

The process of producing and delivering FAIR data (e.g. Wilkinson et al., 2016) is important to develop solutions in order to serve data users as well as data producers (beyond space agency archives, such as NASA PDS or ESA PSA) through a suitable suite of standards, software, and workflows.

On one hand, data producers are busy with collecting ancillary data for their final product that will go public typically through an article describing it; on the other hand, data users are looking for all the available data suitable to support their work. Therefore, the interface between those two categories and how they communicate their results and their demands is of great importance to make the data flow in between as easy as possible.

GMAP [1] develops software, tools, deploying systems and preparing data to cover the gap between producers and consumers of planetary geologic maps (see also Hare et al., 2014, 2018).

Geological maps are a sensible research product to the discussion of data FAIRness for their richness of information (Nass et al., 2021) and great format freedom. Discoverability and Accessibility are covered by data publication platforms such as Zenodo[1], USGS Astropedia[2] (e.g. Bailen et al., 2012), or PANGAEA (Diepenbroek et al., 2002; Devaraju, et al. 2018) by providing easy-to-use interactive, searchable web interfaces to data stores.

The GMAP data integration portal architecture

Within the GMAP geological data integration portal (Rossi et al., 2020, D8.2), a set of services have been set up in order to provide public dataset access to GMAP and the wider community. Based on mapping standards (See Nass et al., 2020; D9.1), preparation of partial products and associated metadata (Rossi et al., 2020; Nass et al., 2020; D9.1, and appendix therein) – described in the preceding deliverables, data are made publicly accessible through the portal.

In the overall idea of FAIRness (see also Brandt et al., 2020), besides the attributes directly associated to the meta/datasets for an optimal use of the data, the software component is equally important, in order to provide an easy, intuitive workflow at the final stage for both data producers/providers (i.e. geologic mappers) and users.

For data publishing, we upgraded the custom-made solution developed at the beginning of the project[3] to a professional, Open Source, industry-standard solution provided by CERN: Invenio Research Data Management (Invenio-RDM). Invenio is the software behind Zenodo, the first production suitable version was released in August 2021 (InvenioRDM v6.0)[4].

Invenio-RDM provides the same interface as Zenodo, which is beneficial to our users' experience. Another important aspect for choosing Invenio was the possibility to customise the metadata of the publication itself, which will allow us to provide custom search capabilities to our planetary data, as well as cross-linking to other geologic map (or basemap) repositories and publication venues.

---

[1] https://www.zenodo.org

[2] https://astrogeology.usgs.gov/astropedia

[3] https://github.com/europlanet-gmap/data_site

[4] https://inveniordm.docs.cern.ch/releases/versions/version-v6.0.0/

Invenio is composed of an app that handles the web-gui (graphical user interface) as well as the programmatic API (application programming interface), available at https://data.europlanet-gmap.eu and https://data.europlanet-gmap.eu/api, respectively. A set of services are deployed around the app to handle the data internals – data archive, database, free-text search – as well as load-balancing on multiple-users requests. Figure 1 presents the whole architecture responsible for the service.
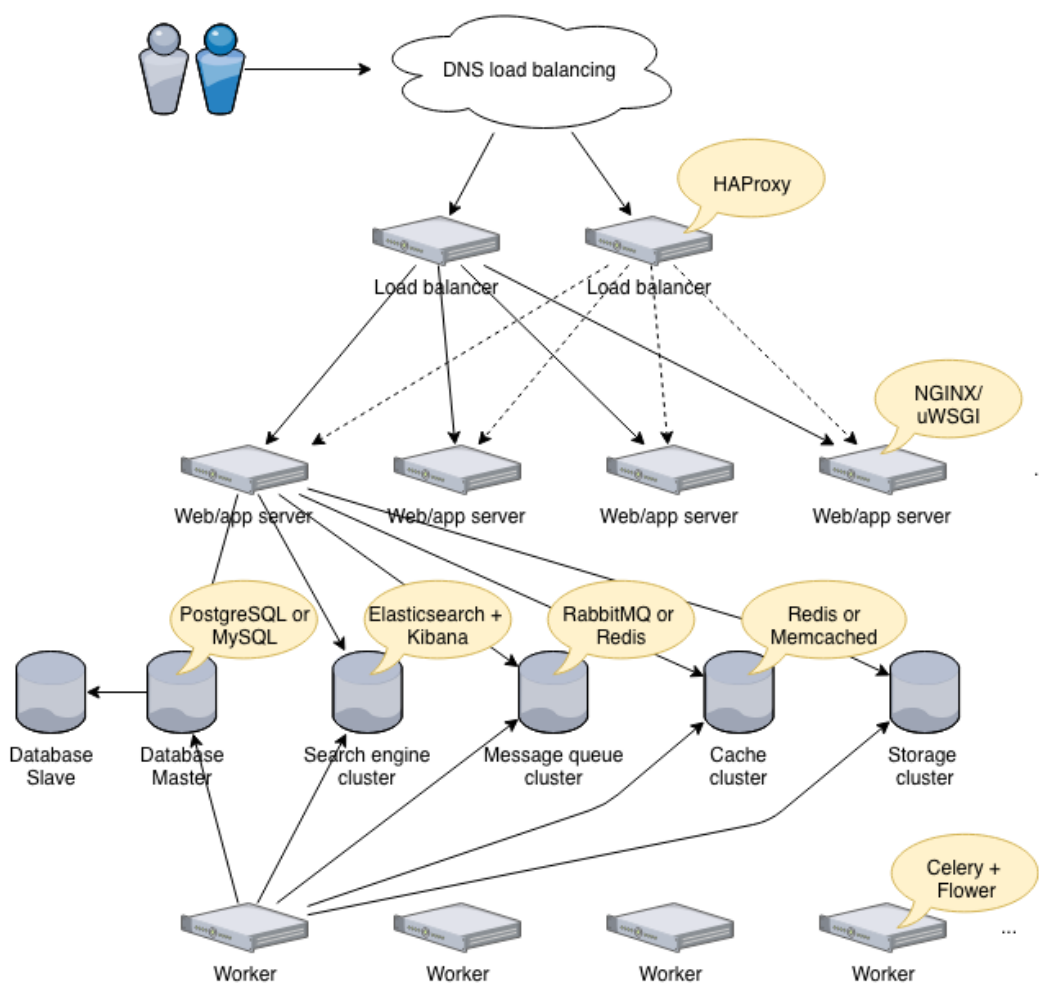

Figure 1: Invenio architecture (Credit: Invenio training documentation). Custom metadata can be added.

The entry page of the data portal (see for comparison with the initial version D8.2) is visible in Figure 2.
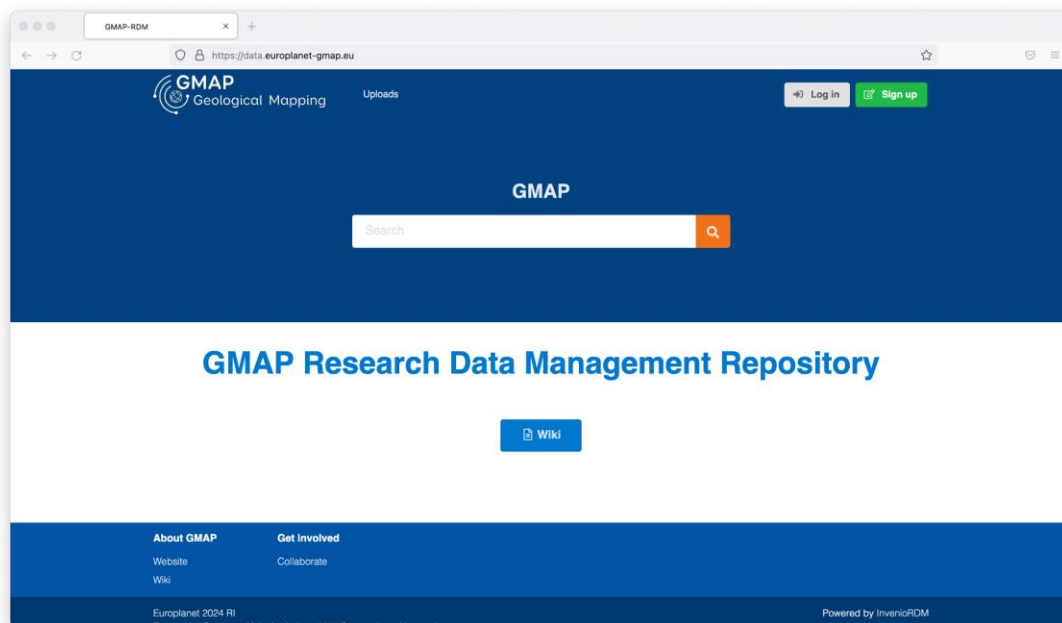
Figure 2: Screenshot of the updated Data portal page (see also Rossi et al,. 2020, D8.2), accessible at https://data.europlanet-gmap.eu/.

Portal workflow

Within the GMAP portal, Invenio is responsible for the project data management of geological maps as well as any ancillary data (e.g., raster basemaps). On GMAP packages, users will use public data and processing facilities – or their preferred tool – to produce geological maps following GMAP guidelines to eventually package the final data product in a `zip` archive and upload into GMAP's Invenio-RDM instance. Nevertheless, individual raster and vector data files, as well as, final geological map documents can also be uploaded by specifying the different content types (dataset, image, table, document).

The main idea, in the workflow for the making of maps, is to have the geological maps and ancillary data right next to each other, available in the same environment under our premises. This is accomplished by the use of Invenio's API interface[5], which allows direct access through programmatic approaches (Python, Shell), provided by the JupyterHub service.

Figure 3 presents the general view of GMAP (software and data) architecture. Invenio shares a common environment with JupyterHub and the Maps web-app, where the published packages as well as ancillary data products can be visualised and accessed through common interfaces.

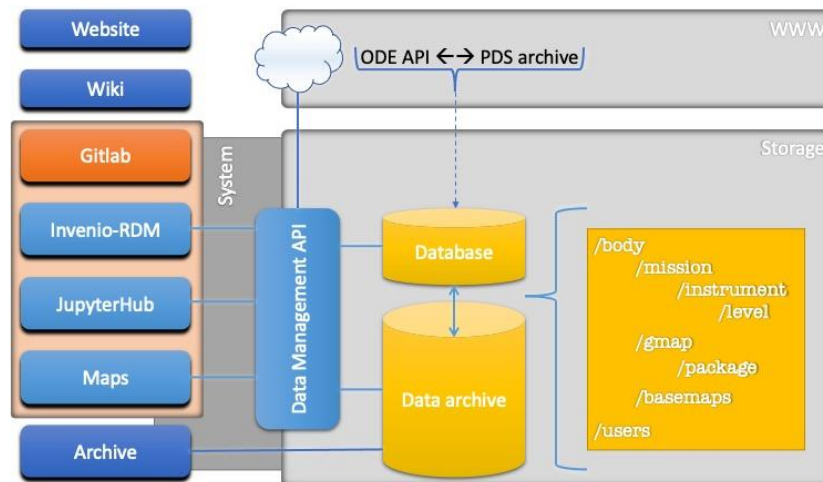[5] https://inveniordm.docs.cern.ch/install/run/

Figure 3: GMAP services architecture.

The envisaged GMAP engagement levels (see Rossi et al., 2022; D8.5) are described below. In bold are listed those with specific reference to the GMAP Data Integration Portal, i.e. directly hosted or linked/discoverable as metadata-only entries (Figure 4):

- Level 0 - Occasional users at the very beginning of their career but curious about planetary geology, casual participants to informal monthly calls, active on social and relaunching GMAP content
- Level 1 - Participants in yearly / periodic schools and workshops, users of GMAP materials, e.g. registered to the GMAP web to access Winter School videos, or GitHub users (e.g. via stars, forks, other measurables) and casual participants to informal monthly calls
- *Level 2 - Geologic mappers with published maps who make them discoverable via the GMAP portal (i.e. maps hosted and published elsewhere, but linked/listed from the portal (no datasets on repo)*
- *Level 3 - Geologic mappers with published papers or other products who share data on GMAP Portal / Zenodo, using the portal but not complying to all standards (dataset hosted by the portal)*
- *Level 4 - Geologic mappers with or without published papers who engage with GMAP and develop their project (from scratch or from a certain point) with GMAP, adhering to standards and joining calls/seminars where relevant/possible (dataset hosted by the portal)*
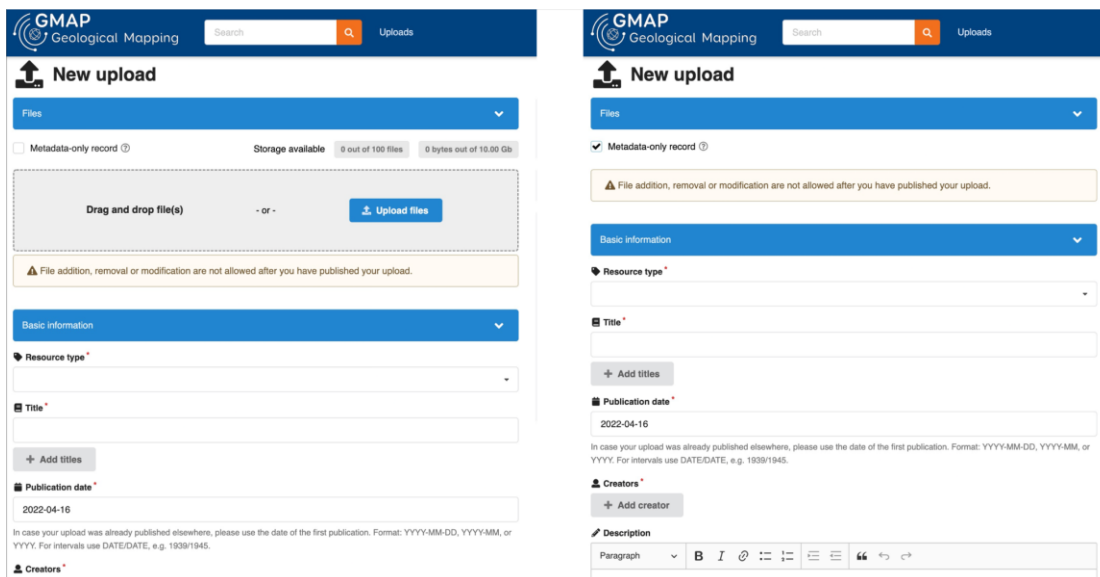
Figure 4: Left: view of the regular data & metadata entries for uploading a dataset (e.g. geospatial data, basemaps, etc.). Right: view of the Invenio/GMAP metadata-only entry addition, suitable for making discoverable geologic maps/basemaps already published on existing 3rd party data repositories.

Data published

Data published so far includes legacy PLANMAP data (See Rossi et al., 2020; D8.2) and published datasets from PLANMAP/GMAP community mapping projects (see Qian, Y., et al. 2020; 2021). As soon as current community mapping projects produce data (see Level 2 to 4 above), they will be hosted or linked from within the portal.
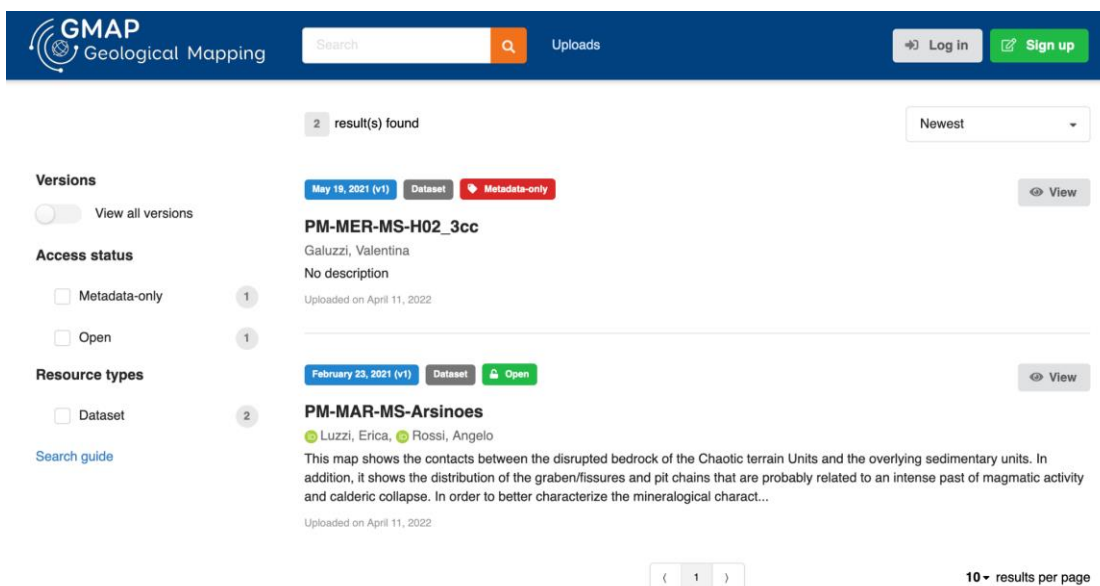

Figure 5: exemplary search results on the GMAP Data Integration Portal[6].

Additionally, some preliminary cross-discovery metadata-only entries from USGS exist on the GMAP Data portal, with more to come. The process of importing and parsing

---

[6] https://data.europlanet-gmap.eu/search?q=&l=list&p=1&s=10&sort=newest

into the GMAP Data Integration Portal USGS (or in principle, other) datasets can be performed programmatically[7].

Outlook and Plans

The backend of the data portal and its ease of use with existing data publication services such as OpenAIRE Zenodo[8] is suitable for making available a variety of GMAP-relevant scientific datasets, manually, programmatically and/or semi-automatically (see also Rossi et al., 2018).

Current plans for the GMAP Data Integration Portal include:

- Homogenisation of metadata and interaction with USGS / Astropedia (see e.g. Hare et al., 2014, 2018)
- Add geo-metadatas for products for easier discovery (also using internal invenio elasticsearch)
- Improve interoperability through GMAP / USGS et al.
- Include VESPA entries of GMAP products for relevant datasets , similarly to PLANMAP[9]

References

Bailen M.S. (2012) Astropedia – A Data Portal for Planetary Science, 43rd Lunar and Planetary Science Conference, https://www.lpi.usra.edu/meetings/lpsc2012/pdf/2478. pdf .

Brandt, C. H., Rossi, A. P., Penasa, L., Pozzobon, R., Luzzi, E., Wright, J., Carli, C., and, Massironi, M. (2020) PLANMAP data packaging: lessons learned towards FAIR planetary geologic maps, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020-18839, https://doi.org/10.5194/egusphere-egu2020-18839

Devaraju A. et al. (2018) PANGAEA Data Recommender: An Out of the Box Approach to Discover Earth and Environmental Datasets, EGU General Assembly 2018.

Diepenbroek M. et al. (2002) PANGAEA – An Information System for Environmental Sciences, Computer & Geosciences, 29(10), https://doi.org/10.1016/S0098-3004(02)00039-0

Hare, T. M., Keszthelyi, L., Gaddis, L., & Kirk, R. L. (2014). Online planetary data and services at USGS astrogeology. In Lunar and Planetary Science Conference (Vol. 45, p. 2487).

---

[7] see https://github.com/europlanet-gmap/invenio_tools

[8] https://zenodo.org/

[9] https://epn1.epn-vespa.jacobs-university.de

Hare, T. M., Rossi, A. P., Frigeri, A., & Marmo, C. (2018). Interoperability in planetary research for geospatial data analysis. Planetary and Space Science, 150, 36-42.

Nass et al., (2020) D9.1 GMAP Standard Definition Document, Europlanet H2020 RI deliverable, available online at: https://wiki.europlanet-gmap.eu/bin/view/Main/Deliverables

Nass A. et al. (2021) Facilitating reuse of planetary spatial research data – Conceptualising an open map repository as part of a Planetary Research Data Infrastructure, Planetary and Space Science (v.204), https://doi.org/10.1016/j.pss.2021.105269.

Qian, Yuqi (2020), "CE5 - Young Mare Basalts", Mendeley Data, V1, doi: 10.17632/5dnt9h58px.1

Qian, Y., Xiao, L., Head, J. W., van der Bogert, C. H., Hiesinger, H., & Wilson, L. (2021). Young lunar mare basalts in the Chang'e-5 sample return region, northern Oceanus Procellarum. Earth and Planetary Science Letters, 555, 116702.

Rossi, A. P., Shin, J., Marco Figuera, R., Minin, M., Manaud, N. (2018) Mapping bibliometrics for Planetary Science, EPSC Abstracts Vol. 12, #EPSC2018-677-1.

Rossi A. P. et al, (2020) D8.2 GMAP Data Integration Portal, Europlanet H2020 RI deliverable, available online at https://wiki.europlanet-gmap.eu/bin/view/Main/Deliverables/

Rossi et al., (2022) D8.5 GMAP Community mapping report 1, Europlanet H2020 RI deliverable, available online at: https://wiki.europlanet-gmap.eu/bin/view/Main/Deliverables/

Wilkinson, M.D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3, https://doi.org/10.1038/sdata.2016.18