DOI: 10.26314/GARR-Conf25-proceedings-16

Collaborative and Reproducible science infrastructure: the Europlanet GMAP JupyterHub processing environment

Giacomo Nodjoumi[1,2,3], Carlos H. Brandt[4], Javier Suárez-Valencia[3], Erica Luzzi[5], Mario Valiante[6], Veronica Camplone[1,2], Edoardo Rognini[1,2], Angelo Pio Rossi[7], M. Giardino[1,8], A. Zinzi[1,8]

[1]Space Science Data Center (SSDC), Agenzia Spaziale Italiana (ASI), Via del Politecnico snc, 00133, Rome, Italy, [2]INAF/Osservatorio Astronomico di Roma (INAF-OAR), Via Frascati 33, 00078, Monte Porzio Catone (RM), [3]School of Science, Constructor University Bremen gGmbH, Bremen, DE, [4]EGI Foundation, Amsterdam, NL, [5]Mississippi Mineral Resources Institute, University of Mississippi, 114 Brevard Hall, University, USA, [6]Dipartimento di Ingegneria Civile, Università degli Studi di Salerno, Fisciano, Italy, [7]Earthgraph GmbH, Bremen, Germany, [8]Agenzia Spaziale Italiana (ASI) - Via del Politecnico snc, 00133, Rome, Italy

Abstract. Scientific research frequently uses advanced computational tools and skills to handle its large datasets. We developed a user-friendly, innovative solution based on standardized Dodgeville recipes and ipykernels to streamline interactive data processing environment deployment and management. Our solution is compatible with a series of widely used analytical software based on Python, and R, and specialized software for planetary sciences such as the USGS Integrated Software for Imager and Spectrometers (ISIS) and the NASA Ames Stereo Pipeline (ASP), providing a scalable and highly customizable solution for collaborative research, teaching, and specialized group.

Introduction

Research data are available in various formats, ranging from raw measurements to processed and calibrated products. Accessing such data is facilitated by dedicated archives and platforms such as NASA's Planetary Data System (PDS) (PDS Geosciences Nodes, 2020), ESA's Planetary Science Archive (PSA) (Planetary Science Archive, 2025) , Mars System of Information (MarSI) (Quantin-Nataf et al., 2018), SSDC-ASI's Multi-purpose Advanced Tool for Instruments of the Solar System Exploration (MATISSE) (Zinzi et al., 2016), ESA MAAP (Multi-Mission Algorithm and Analysis Platform) (Albinet et al., 2019) and general-purpose environments like Google Colab. However, turning these data into useful scientific knowledge usually needs special software and expertise in computing. Several initiatives aim to provide Analysis Ready Data (ARD) (Building a Lunar Spatial Data Infrastructure (SDI) - ADS, 2021; Knowledge Inventory of Foundational Data Products in

Conferenza GARR 2025 - Frontiere Digitali - Selected Papers
Bari 13-15 maggio 2025 - ISBN: 978-88-946629-4-8

Planetary Science, 2021; Fergason et al., 2021) and interactive web-based analytical interfaces [10]. However, many advanced analyses still depend on sophisticated, open-source software packages such as the USGS Astrogeology Research Program's ISIS and (Sucharski et al., 2020) NASA Ames Stereo Pipeline (ASP) (Beyer et al., 2018). Despite their effectiveness, these tools require familiarity with UNIX systems and programming, presenting a barrier to broader adoption among researchers.

Proposed Solution

To address these challenges, we developed a versatile, scalable solution using Docker containers (Forde et al., 2018) and JupyterHub (Kluyver et al., 2016). Our solution, namely Europlanet GMAP JupyterHub (Nodjoumi et al., 2025), emphasizes standardized Dockerfile recipes enhanced with ipykernels, facilitating easy adaptation for specific tasks or research working groups. The core image is the ISIS-ASP-GISPY Docker image, which integrates ISIS, ASP, and a curated collection of Python packages (GISPY - Geospatial Python), allowing customization for diverse analytical requirements. Docker containers offer a lightweight, conflict-free environment, while JupyterHub provides a web-based platform for interactive computing through notebooks that integrate code, output, and explanatory text. A schematic of the architecture is presented in Figure 1.

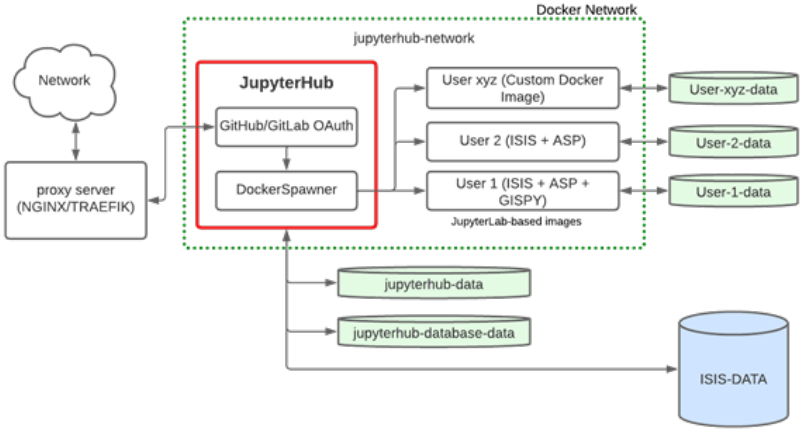


Fig. 1 Schematic of the proposed architecture. Boxes represent docker containers, while cylindres represent docker volumes.

Deployment and Management

The system is deployed through a semi-automated script handling Docker volume creation, networking, and image configuration. Additional parameters enable advanced settings for web service deployment, user access control, and resource allocation. Researchers interact with their computing environments through an accessible JupyterHub web interface, simplifying adoption and use. Shared folders for users can be configured as well as third-party services connections, such as GitHub/GitLab and cloud storages like Google Cloud.

Advantages and Impact

Our method presents substantial benefits over traditional computational workflows:

- Simplified Deployment and Maintenance: Dockerized environments reduce complexity and administrative burden.
- User-Friendly Interface: JupyterHub provides intuitive access, catering to various levels of technical expertise.
- Scalability: Easily adaptable for expanding user groups and/or workloads, suitable for collaborative research, educational settings, and workshops.
- Customization: The environment and software packages can be tailored to specific research needs, accommodating additional tools as required.

For instance, curated Docker images can be used on a laptop as standalone containers for development and then the developed code can be executed in an identical Docker container running on HPC nodes running JupyterHub without reproducibility issues.
Another example is a data centre composed of several teams, each with its own specialized docker image, customized with multiple ipykernels, each for specific task.
Initially developed within the Europlanet GMAP (Geological Mapping of Planetary Bodies) project at Constructor University, this solution was tested at international conferences such as COSPAR. Currently, it is being implemented at the Space System Data Center (SSDC) of the Italian Space Agency (ASI). After an initial assessment and integration with the other SSDC tools and services, we plan to gradually open and enable the JupyterHub service to the wider community of researchers and citizens. Final users will then have access to a user-friendly environment, remotely accessible, with embedded access to data, tools and pipelines running on high-end HPC, relieving the user to perform extensive computing tasks on personal computers and laptops.
A semi-production-ready implementation is publicly available on Zenodo and GitHub (Brandt et al., 2024). Continuous improvements are underway to enhance user experience, particularly focusing on session persistence and recovery functionalities critical for lengthy computational tasks. Further details about GMAP can be found at https://www.europlanet.org/.

Conclusion

This standardized and customizable computational environment significantly simplifies access to advanced analytical tools, promoting broader participation across diverse research communities. By lowering technical barriers, it supports collaborative, scalable, and effective data analysis, ultimately facilitating scientific analyses and innovation. Moreover, the HPC-based infrastructure, as well as the integration and accessibility of multiple datasets, services and tools already available at the SSDC, may also allow the application of novel processing technologies on even very large datasets and/or old datasets, further improving and widening the scientific outcome.

References

Albinet, C., Whitehurst, A. S., Jewell, L. A., Bugbee, K., Laur, H., Murphy, K. J., Frommknecht, B., Scipal, K., Costa, G., Jai, B., Ramachandran, R., Lavalle, M., & Duncanson, L. (2019). A Joint ESA-NASA Multi-mission Algorithm and Analysis Platform (MAAP) for Biomass, NISAR, and GEDI. Surveys in Geophysics, 40(4), 1017–1027. https://doi.org/10.1007/s10712-019-09541-z

Beyer, R. A., Alexandrov, O., & McMichael, S. (2018). The Ames Stereo Pipeline: NASA's Open Source Software for Deriving and Processing Terrain Data. Earth and Space Science, 5(9), 537–548. https://doi.org/10.1029/2018EA000409

Brandt, C. H., Nodjoumi, G., Rossi, A. P., Luzzi, E., & Suarez Valencia, J. (2024). europlanet-gmap/docker-jupyterhub: First Release (Versione 1.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.14555694

Building a Lunar Spatial Data Infrastructure (SDI)—ADS. (2021). https://ui.adsabs.harvard.edu/abs/2021LPICo2549.7054H/abstract

Fergason, R. L., Hunter, M. A., Laura, J. R., & Hare, T. M. (2021). Analysis Ready Data Available Through the SpatioTemporal Asset Catalog (STAC) Specification: Investigating the Application to Planetary Data. 2549, 7023. https://ui.adsabs.harvard.edu/abs/2021LPICo2549.7023F

Forde, J., Head, T., Holdgraf, C., Panda, Y., Nalvarete, G., Ragan-Kelley, B., & Sundell, E. (2018). Reproducible Research Environments with Repo2Docker. https://openreview.net/forum?id=B1lYOwuoxm

Kluyver, T., Ragan-Kelley, B., P&#233, Rez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., n, Abdalla, S., Willing, C., & Team, J. D. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas (pp. 87–90). IOS Press. https://doi.org/10.3233/978-1-61499-649-1-87

Knowledge Inventory of Foundational Data Products in Planetary Science. (2021). https://essopenarchive.org/doi/full/10.1002/essoar.10501479.1

Nodjoumi, G., Brandt, C. H., Suárez-Valencia, J. E., Luzzi, E., Valiante, M., & Rossi, A. P. (2025). Collaborative and Reproducible Planetary Science Through the Europlanet GMAP JupyterHub Processing Environment. Earth and Space Science, 12(5), e2025EA004251. https://doi.org/10.1029/2025EA004251

PDS Geosciences Nodes. (2020). PDS Geosciences Node Orbital Data Explorer (ODE). https://ode.rsl.wustl.edu/

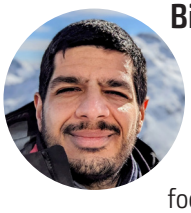
Planetary Science Archive. (2025). https://psa.esa.int/psa/#/pages/home

Quantin-Nataf, C., Lozac'h, L., Thollot, P., Loizeau, D., Bultel, B., Fernando, J., Allemand, P., Dubuffet, F., Poulet, F., Ody, A., Clenet, H., Leyrat, C., & Harrisson, S. (2018). MarsSI: Martian surface data processing information system. Planetary and Space Science, 150, 157–170. https://doi.org/10.1016/j.pss.2017.09.014

Sucharski, T., Mapel, J., Jcwbacker, Kristin, Lee, K., AgoinsUSGS, Shepherd, M., Combs, C. R., Stapleton, S., Dcookastro, Rodriguez, K., Becker, K. J., Sides, S., Cole, Jusflag, Wilson, T., Acpaquette, Williams, K., Jlaura, … Rsaleh57. (2020, luglio). USGS-Astrogeo-

logy/ISIS3: ISIS 4.2.0 Public Release (Versione 4.2.0). Zenodo. https://doi.org/10.5281/zenodo.3962369
Zinzi, A., Capria, M. T., Palomba, E., Giommi, P., & Antonelli, L. A. (2016). MATISSE: A novel tool to access, visualize and analyse data from planetary exploration missions. Astronomy and Computing, 15, 16–28. https://doi.org/10.1016/j.ascom.2016.02.006

Bibliography



Giacomo Nodjoumi

Giacomo Nodjoumi holds a Bachelor's degree in Geology and a Master's degree in Engineering Geology, Land Use Management and Georisks, both from Sapienza University of Rome. He earned a PhD in Geosciences from Constructor University Bremen, with a thesis focused on the automatic detection of pits, skylights, and cave candidates on the Moon and Mars using remote sensing and deep learning techniques and subsurface sounding radar data analyses. He is currently a research fellow at the Italian Space Agency's Space Science Data Center (ASI-SSDC), where he works on the development and deployment of scientific data services, such as the Europlanet GMAP JupyterHub, and other services on HPC systems. He contributes to several European projects, including the Horizon 2020 EXPLORE project, where he helped design tools like L-EXPLO and L-HEX for lunar data visualization and analysis, and Europlanet's GMAP initiative, supporting the creation of reproducible workflows for planetary surface mapping.

Carlos Brandt

Carlos Brandt is a Software Architect at the EGI Foundation, where he is part of the Technical Solutions team. With extensive expertise across multiple layers of the scientific software stack — including system administration, database modeling, numerical simulation, image processing, and distributed computing — Carlos brings a comprehensive approach to scientific computing and infrastructure design.
He holds a PhD in Astrophysics from the Sapienza University of Rome, building on a background that includes a degree in Physics from the Federal University of Rio Grande do Sul and a Master's in Numerical Simulations from the National Laboratory for Scientific Computing, both in Brazil. His academic and research trajectory led to roles in computational astrophysics at the Italian Space Agency's Data Center and as a lecturer in Data Engineering at Jacobs University Bremen.
Now based in Germany, Carlos has contributed to several European projects focused on space and geospatial data engineering. He is a strong advocate for open-source software and is passionate about initiatives that promote open access to scientific knowledge.

Erica Luzzi

Dr. Erica Luzzi is a planetary geologist specializing in Mars surface processes, ice detection, and terrestrial analogues. With international experience working at NASA, participating to ESA campaigns, and working in several universities across Europe and the US, she contributed to field campaigns in extreme environments and published widely on Martian geology, analogs for Enceladus, and resource mapping for future human exploration missions.

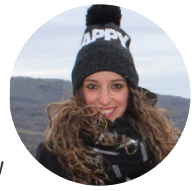
Javier Suaréz-Valencia

Geologist from the National University of Colombia and Ph.D. from Constructor University

Bremen in Germany, currently a postdoc in the University of Padova. He has worked for over eight years studying planetary surfaces using remote sensing techniques, mainly on Earth, Mars, and Pluto. Currently, for his postdoctoral project, he is analysing the geomorphology and composition lava tubes as potential analogues for lunar exploration missions.

Mario Valiante

Mario Valiante is a Research Fellow at the Department of Civil Engineering of the University of Salerno (Academic Discipline GEO/04 – Physical Geography and Geomorphology) since January 2022, where he also serves as lecturer in General Geology and Geomorphology. He obtained his Master's degree with honors in Engineering Geology, Land Use Management and Georisks from Sapienza University of Rome in 2015. In 2016, he qualified as a Professional Geologist (Regional Register of Campania, Section A).
In 2020, he earned a PhD in Earth Sciences from Sapienza University of Rome. In 2021, he held a postdoctoral research contract at the Department of Civil Engineering of the University of Salerno, working across the disciplines GEO/04 and GEO/05.
His main research interests focus on the analysis of landslide phenomena in geologically complex environments and on GIScience, particularly the development of data structures for geospatial information related to geological risk.



Veronica Camplone

Veronica Camplone is a geologist (PhD. in Earth and Environmental Sciences) specializing in remote sensing of Earth and planetary surfaces, with a focus on sedimentology and geomorphology. She also holds a postgraduate degree in Space Institutions and Policies, combining scientific and institutional expertise. Since 2021, she has been a research fellow at INAF in Rome, where she works on planetary data analysis and contributes to the development of the MATISSE tool for visualizing and comparing data from space missions.

Edoardo Rognini

Edoardo Rognini graduated in Physics with a specialization in Astronomy and Astrophysics and obtained a PhD working on radiative diffusion and levitation in low-mass stars. Since 2017, he has been living and working in Rome (initially at INAF-IAPS, then at ASI-SSDC), focusing on thermophysical modeling of airless bodies and thermal data analysis from space missions in the Solar System.

Angelo Pio Rossi

Angelo Pio Rossi is a planetary geoscientist, practitioner, and enterprenuer with 20+ experience in designing, implementing projects/programs on: Geoscience research, Remote Sensing, geospatial data handling, planetary mapping, training and education. More info on http://aprossi.eu and http://earthgraph.eu/

Marco Giardino

Marco Giardino holds a Master Degree in Computer Engineering. His main professional interests are software engineering and scientific data management, with a strong focus on FAIR principles and open science.

He has taken part in the scientific space missions Mars Express, Dawn, and ExoMars and has been involved in several others. He is currently working at the Italian Space Agency, where he leads the IT activities of the Space Science Data Center.

Angelo Zinzi

Angelo Zinzi has both Master Degree and PhD in Physics, with topics relevant to Planetary Sciences. He is now a staff technologies at the Italian Space Agency (ASI) and his main aims are comprised in the management of data of planetary exploration missions using FAIR principles and international standards, such as Virtual Observatory, and he developed the scientific webtool MATISSE. He has been or is involved in a series of space missions, such as ESA Rosetta, ASI LICIACube, ESA JUICE and participated to a series of international projects, such as NEOROCKS.