Astronomical Data Analysis Software and Systems XXX
ASP Conference Series, Vol. 532
J. E. Ruiz, F. Pierfederici, and P. Teuben, eds.

Scientific Data Applications for Science Analysis Platforms

Nick Cox and Jeronimo Bernard-Salas
ACRI-ST, 260 Route du Pin Montard, Sophia-Antipolis, France;
nick.cox@acri-st.fr

Abstract. Science analysis platforms constitute a new paradigm in the exploration and exploitation of existing and future space science data. The advent of big data in space sciences is necessitating researchers to move to the data rather than to move the data to them. The success of these science platforms, which are linked to science archives, relies in part on the quality of the provided generic and domain specific services and tools, and the ease with which users can integrate them into their scientific workflow. These new platforms also offer clear ways on how to ensure collaborative, reproducible and open science and to adopting the FAIR data principles. In this contribution we highlight several new initiatives to provide scientific data applications for science analysis platforms.

1. Introduction

Advances in big data and cloud computing are leading to new ways for scientists to approach their research, impacting directly how science progresses and discoveries are made. Equally, advances in technology are leading to a new data deluge in most scientific domains. A new trend is to move data storage, processing, analysis, and visualisation away from the limited individual workstation to more powerful (federated) computational infrastructures. Furthermore, open data and open science have been acknowledged as key enablers to spur innovation and technological progress for the benefit of society.

Astronomy and (particle) astrophysics' new facilities are poised to generate, already this decade, data on the Petabyte scale and beyond. There is a growing demand for tools and services to keep abreast of this wave of data. In this contribution, in the context of existing and future science analysis platforms, we highlight several recent initiatives geared towards the development of scientific data applications for astronomy and planetary science.

2. Science Analysis Platforms for Space Sciences

Science (analysis) platforms (SAPs) are the logical extension to science archives which centre on discovery and (limited) visualisation. These are web-based services that allow users (researchers, citizen scientists) to discover, visualise, analyse, and (collaboratively) interact with scientific data.

Two prominent European initiatives in the space sciences[1] domain are the ESCAPE Science Analysis Platform (ESAP) and the ESA Datalabs (EDL) both discussed elsewhere in this volume (Allen et al. 2021; Dickinson et al. 2021; Navarro 2021). Both platforms aim to provide their communities with novel tools to discover and analyse their own (and those of others) data resources, generated by large research facilities and space missions, and which are becoming increasingly large and complex. Science platforms are becoming thus an important part of the scientific workflow, offering generic coding applications as well as more domain-specific ones. They give researchers and developers a unique opportunity to deploy their scientific applications on a powerful computing infrastructure. Collaborative research is another key element. Data, code, results can be shared both privately, within a research team, and publicly for dissemination.

It is planned that such science platforms, deployed on centralised or federated cyberinfrastructures, will operate also as part of the European Open Science Cloud (EOSC; Molinaro et al. 2019) which aims to provide access to all research data in a systematic, transparent and sustainable way through a federated approach for data and service providers. Science platforms can thus promote "open science" and support application of FAIR principles to data and software.

The uptake of science platforms depends on the diversity and usefulness of the offered services, a key aspect of which are the science data applications (SDAs). A successful example of a science platform (for life sciences in this case) is Cyverse,[2] which has some 90K users and has already integrated hundreds of (user provided) applications (Merchant et al. 2016).

Next we highlight some ongoing activities, that ACRI-ST[3] is participating in and leading, for the creation of SDAs for the scientific exploitation of space- and ground-based astrophysics and planetary science data. These will enhance the value of science platforms and will also serve as examples and provide templates to help the community bring their own applications to these platforms.

3. Scientific Data Applications for Science Platforms

The following applications (SDAs) are being developed together with scientists and engineers from public institutions and the private sector and leveraging artificial intelligence and human intelligence to provide collaborative, cloud-enabled thematic science applications for the space science community.

3.1. Mars atmospheric exploration service

ACRI-ST is working on a science app to explore and visualise Mars EXpresss/SPICAM atmospheric UV and IR sounding data. These data include vertical density profiles of atmospheric gases (H_2O, O_2, CO_2) and properties of aerosol particles. The application (web service) will allow to interactively query the high-level processed SPICAM data

[1] Space sciences are understood to include all domains related to the exploration of the universe, from solar and planetary science, astronomy to (particle) astrophysics.

[2] `https://www.cyverse.org`

[3] `https://astro.acri-st.fr`

products from the ESA Planetary Science Archive (PSA), visualise the atmospheric retrieval results as well as the underlying transmission or reflection spectra, and subsequently apply processors for data manipulation and analysis (e.g. to create maps, data cubes). Other similar datasets (e.g. from ExoMars Trace Gas Orbiter) may be added in the future. The application can be ported to explore Venus EXpress/SPICAV high-level data products.

3.2. Detection of surface features on Mars

As member of the machine learning (ML) joint research activity (JRA) of the Europlanet Research Infrastructure 2024[4] ACRI-ST is participating to science cases on the automated detection of geological features, such as pits, on the martian surface using state-of-the-art deep learning algorithms. The science cases addressed by this JRA aim to promote and stimulate the uptake of ML for the planetary sciences. We are also working to bring the developed tools to the planetary science community through a dedicated ML Portal[5] and other (online) venues.

3.3. Lunar and Gaia apps

EXPLORE (Innovative Scientific Data Exploration and Exploitation Applications for Space Sciences) is a Horizon 2020 project (2020-2023) whose main objective is to further the scientific exploitation of space science data. This will be achieved by developing a set of six SDAs that are to be deployed on different SAPs. The SDAs cover a broad range of topics, from galactic archaeology, 3d-mapping of interstellar gas and dust, stellar spectral energy distributions, AI based discovery in massive spectral data sets to lunar surface mapping and lunar exploration. Design and development of each SDA is led by scientific experts supported by AI and visualisation experts and software engineers. EXPLORE SDAs will be first deployed, verified and validated on a bespoke science thematic exploitation platform[6].

3.4. Data fusion for JWST

In preparation for the James Webb Space Telescope (JWST), and as part of the ANR LabCom initiative, a new "common laboratory" has been created partnering ACRI-ST and IAS (Université Paris-Sud and CNRS). This lab has the goal to research and innovate novel data fusion methodologies (Bayesian inversion and AI based) to be applied to the JWST/MIRI instrument as well as to remote sensing imagery obtained with the Sentinel-2 and Sentinel-3 optical instruments. The goal for JWST is to fuse the large field-of-view broad-band imaging data with the small field-of-view high-resolution spectral data.

Acknowledgments. The EXPLORE project has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 101004214. Europlanet RI 2024 has received funding from the European Union's Horizon 2020 research and innovation programme under agreement No 871149.

---

[4] `https://www.europlanet-society.org/europlanet-2024-ri`

[5] `https://ml-portal.oeaw.ac.at`

[6] `https://explore-platform.eu`

References

Allen, M. G., Lamanna, G., Espinal, X., Graf, K., van Haarlem, M., Serjeant, S., Bird, I., Cuoco, E., & Wagh, J. 2021, in ADASS XXX, edited by J. E. Ruiz, F. Pierfederici, & P. Teuben (San Francisco: ASP), ASP Conf. Ser.

Dickinson, H., Vermaas, N., & Meyer-Zhao, Z. 2021, in ADASS XXX, edited by J. E. Ruiz, F. Pierfederici, & P. Teuben (San Francisco: ASP), ASP Conf. Ser.

Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., & Antin, P. 2016, PLOS Biology, 14, 1. URL https://doi.org/10.1371/journal.pbio.1002342

Molinaro, M., Allen, M., Bertocco, S., Boisson, C., Bonnarel, F., Castro Neves, M., Demleitner, M., Genova, F., Morris, D., Schaaff, A., Taffoni, G., & Voutsinas, S. 2019, arXiv e-prints, arXiv:1911.08205. 1911.08205

Navarro, V. 2021, in ADASS XXX, edited by J. E. Ruiz, F. Pierfederici, & P. Teuben (San Francisco: ASP), ASP Conf. Ser.