SF2A 2023
M. N'Diaye, A. Siebert, N. Lagarde, O. Venot, K. Baillié, M. Béthermin, E. Lagadec, J. Malzac, J. Richard (eds)

ASTRONOMY AND THE EUROPEAN OPEN SCIENCE CLOUD

M. Allen[1], B. Cecconi[2], M. Marchand[1] and A. Gonneau[1]

Abstract. The European Open Science Cloud (EOSC) is an initiative to create a 'web of FAIR data' to enable open science. We present an overview of the astronomy-related activities that are contributing to the development of the EOSC. We highlight the participation of the Centre de Données astronomiques de Strasbourg (CDS) and the Observatoire de Paris in EOSC projects, and provide examples of how EOSC is already being used to support astronomy research and data sharing.

Keywords: Astronomical data bases, Virtual observatory tools, Gravitational waves, Standards, Open Science, Data sharing

1 Introduction

The vision of the European Open Science Cloud (EOSC) is to set up a 'web of FAIR data and services' to serve science in Europe. The first steps toward EOSC started in $\sim$ 2018 and current information about the development of EOSC is available on the EOSC Association web pages (`https://eosc.eu/`). The 'Strategic Research and Innovation Agenda' document available there, provides a detailed description of Open Science, the creation of the EOSC concept and its guiding principles, and it also identifies the challenges for its implementation and the potential impacts. The 'Multi-Annual Roadmap' defines a set of priorities for future investment in EOSC, with specific objectives.

Many consultations at European, national and institutional levels have contributed to the overall vision and development of EOSC. Phase 1 of EOSC (2018-2020) was a preparatory period, we are now in a 'convergence phase' (2021-2027), with the 'operational phase' expected to begin in 2027. In this article we summarise the astronomy-related activities that have been pursued at the Centre de Données astronomiques de Strasbourg (CDS*) and the Observatoire de Paris in the context of EOSC, highlighting examples of open science in astronomy and the use of EOSC services. The Action Spécifique Observatoires Virtuels (ASOV) also plays an important role of bringing the French actions toward the Virtual Observatory together, and information about the development of EOSC has been shared at ASOV events.

2 European projects supporting open science in astronomy

The CDS and the Observatoire de Paris are contributing to the development of Open Science in astronomy. The CDS is a Research Infrastructure providing services for reference data in astronomy. CDS participates in many initiatives for the promotion of Open Science at the national, european and international level. CDS recently became a Thematic Reference Centre of the French Recherche Data Gouv initiative, it leads many European Virtual Observatory projects and it is a major actor in the International Virtual Alliance (IVOA). CDS participates in EOSC via the EOSC Association memberships of CNRS and the Université de Strasbourg, and the CDS participation in European projects related to EOSC. The Observatoire de Paris is a member of the EOSC association since its inception. Observatoire de Paris teams are also contributing to international collaborations including similar Open Science and cloud-oriented developments, especially in Japan and in the USA. The development of EOSC and the associated projects is considered by Observatoire de Paris as strategic

[1] Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France
[2] LESIA, Observatoire de Paris-PSL, CNRS, Sorbonne Université, Université Paris Cité, UMR 8109, F-92190 Meudon, France
*`https://cds.unistra.fr`

for upcoming large scale research projects of the institution and the astrophysics science community. As a result, both the CDS and Observatoire de Paris contribute to the shaping of EOSC, at the national level through the *Collège EOSC*, grouping all the French EOSC members, under the umbrella of the French Ministry of Research, the European level of EOSC itself, and also the promotion of European activities at the international level.

2.1 ESCAPE

The European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructure project (ESCAPE) addressed the Open Science challenges shared by ESFRI facilities. These included the astronomy and particle physics infrastructures: SKA, CTA, KM3Net, EST, ELT, HL-LHC, EGO/Virgo and FAIR as well as other pan-European research infrastructures (CERN, ESO, and JIVE) (Allen et al. 2022a). ESCAPE actions were focused on developing solutions for the large data sets handled by the ESFRI facilities. The CDS led a major work package of ESCAPE to "Connect ESFRI projects to EOSC through VO framework" (CEVO), with 16 partners including the astrophysics ESFRI and partners with expertise in the Virtual Observatory (CDS, University of Edinburgh, University of Heidelberg, INTA, INAF, Observatoire de Paris). The main task of this work was the implementation of the FAIR *'findable, accessible, interoperable, re-usable'* principles for ESFRI data through the Virtual Observatory, which was pursued by advancing the definition and adoption of common open IVOA standards for interoperability based on scientific requirements of the ESFRI and the astronomy community (Molinaro et al. 2021; Allen et al. 2022b). The results are described in the deliverable reports[†] of the CEVO work package.

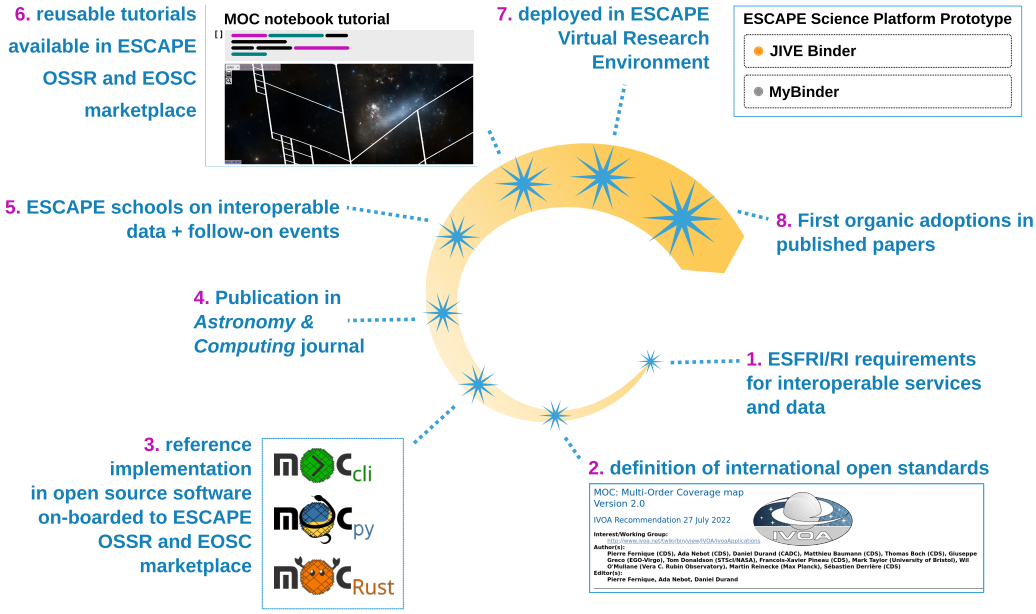

Fig. 1. The Open Science Cycle example of the EGO/Virgo science case in ESCAPE

Here we illustrate the results of the ESCAPE work with the example of the 'open science cycle' developed for the EGO-Virgo requirements. The cycle includes 8 steps from scientific requirements through re-usable notebooks fully integrated into EOSC, and then to community adoption:

1. Requirements for interoperable sky-area and temporal indexation were established, in particlar for managing sky areas in follow-up observations of gravitational wave events.

2. Standardisation was achieved with the acceptance of the MOC 2.0 IVOA recommendation (Fernique et al. 2022) following active participation of the partners in the IVOA.

3. The reference implementation of the MOC standard designed as a software suite in different languages – a command line interface MOCcli, a python library MOCpy (Baumann et al. 2023) and a Rust crate

[†]https://wiki.escape2020.de/index.php/WP4_-_CEVO

MOClibrust (Pineau & Baumann 2021) – to adapt to different users. This led to the release of open source software libraries which have been incorporated into visualisation tools (Baumann & Boch 2023), and on-boarded to the ESCAPE software repository (OSSR) and in the EOSC marketplace.

4. An Open Access refereed publication (Greco et al. 2022) disseminated the technical and scientific results generated by the new standards, tools and application.
5. A tutorial in the ESCAPE training event was based on the methods used for the publication, providing a real science case tutorial.
6. A Jupyter notebook implementation of the tutorial was created and was included into a collection of notebooks that has been on-boarded as a resource to the OSSR Marchand et al. (2023) and as a part of a tutorial suite directly included in the EOSC list of tutorials (Access and use of astronomy-related data from Python: a series of Jupyter notebooks tutorials).
7. Notebook tutorials are compatible with diverse deployments like the Binder interface implemented in ESCAPE Virtual Research Environment or *'devcontainers'* and Jupyter hub as used in the EOSC capabilities.
8. As a result, the first scientific papers explicitly referencing the MOC standard or one of the reference implementation software start to appear, some directly linked to gravitation wave follow-up science Chaudhary et al. (2023), and others from different astronomical domains, e.g. Wang et al. (2023); Rankine et al. (2023); Sweijen et al. (2023).

2.2 EOSC Pillar

The EOSC Pillar project (`https://www.eosc-pillar.eu`) conducted a program around the topic of building EOSC from national contributions. One of the astronomy-related activities that was selected as a thematic use-case was the on-boarding of the CDS SIMBAD service to EOSC. The support of the EOSC Pillar project was important for establishing the CDS as a 'service provider' in the EOSC Portal. The CDS SIMBAD service was then 'on-boarded' so that the service is made findable in the EOSC Portal[‡].

2.3 EOSC Future

The EOSC Future project (`https://eoscfuture.eu`)brought together the e-Infrastructures and Science Cluster communities to build a first implementation of EOSC. One of the astronomy activities in the project led by the CDS was to integrate the documentation about the CDS VizieR publishing process into the EOSC training materials. A tutorial[§] outlining the steps of publishing the data connected with refereed astronomy journal articles was prepared, and on-boarded to EOSC. This tutorial demonstrates how the already established CDS processes for publishing data in astronomy journals leads to the records being visible in the EOSC Marketplace without additional actions by the authors. This work consolidates the ESCAPE work of on-boarding the Virtual Observatory registry to EOSC, and Fig 2 shows the different paths for journal papers, VO services and general services can take to become visible in the EOSC system. Furthermore, a set of python notebook tutorials[¶] based on the ESCAPE 'Science with interoperable data schools' have also been made sustainable (Marchand et al. 2022) and on-boarded as part of the EOSC Future project, see also Fig 1.

3 Using EOSC in Astronomy projects

The main entry point of EOSC, when looking for resources is the EOSC Marketplace[‖], which is gathering all resources available in EOSC (services, data, publications...). A service provider interface is also available to register their services. The activities have focused on the five types of "generic services", which are basic bricks to be used to implement distributed digital infrastructures: (i) *Cloud Computing*, (ii) *Cloud Storage*, (iii) *Persistent Identifiers* and (iv) *Authentication and authorization infrastructure (AAI).*

[‡] `https://search.marketplace.eosc-portal.eu/search/service?q=SIMBAD`
[§] `https://search.marketplace.eosc-portal.eu/trainings/eosc.cds.6a0826b29f1f93ceed88ecf0de4e764c`
[¶] `https://search.marketplace.eosc-portal.eu/trainings/eosc.cds.0d5c7d6448cae721c84e6884f19f5f0b`
[‖] `https://marketplace.eosc-portal.eu`

Fig. 2. Paths for astronomy resources to become visible in the EOSC Marketplace/Portal. i) Published papers with associated data in the CDS VizieR service become visible in EOSC via the successive harvesting of the metadata by the Virtual Observatory registry, the EUDAT B2FIND service, and the integration of B2FIND in EOSC. ii) Similarly Virtual Observatory services in general are harvested by EUDAT B2FIND and hence become visible in EOSC. iii) There is also the possibility of directly on-boarding services, such as has been done for the CDS SIMBAD, with the CDS as the EOSC 'service provider'.

Several use cases have been explored by teams at the Observatoire de Paris, through various projects, namely, ESCAPE, Europlanet, FAIR-IMPACT, NenuFAR and VAMDC:

- The institution was a beneficiary of the ESCAPE project, and thus participated to the developments, through its contribution to the CTA (Cherenkov Telescope Array) project. The NenuFAR team was invited to many ESCAPE meetings for their expertise in low frequency radioastronomy, providing in-kind contribution. Currently, the Observatoire de Paris team has set up a prototype based on ESCAPE developments to run Jupyter notebooks on the local computing cluster. This prototype is aiming at providing a working and training infrastructure for teams willing to develop and deploy their pipeline is an environment, which can be scaled up to EOSC.

- Within the last Europlanet project (EPN2024RI, `https://doi.org/10.3030/871149`), the VESPA (Virtual European Solar and Planetary Access, Erard et al. 2020) team, lead by Observatoire de Paris, has been exploring EOSC services, thanks to the support of the EOSC-Hub project. Their VESPA-Cloud project (Cecconi et al. 2020) has been selected as an "Early Adopter Program", and received direct support from EGI for set up virtual machines on several cloud computing environments available through EOSC (namely: CC-IN2P3 and CESNET). The cloud computing virtual machines are used during the VESPA providers training workshop, and the team is consider using the virtual server in a production context soon. VESPA is also using the Eduteams AAI (`https://eduteams.org`, provided by GÉANT) to provide access to their gitlab server.

- The Observatoire de Paris is a funded participant of the FAIR-IMPACT project (`https://doi.org/10.3030/101057344`). In this project, the team at Observatoire de Paris is prototyping the use of ontologies in astronomy to enhance the semantic interoperability within EOSC.

- The NenuFAR (New Extension in Nançay Upgrading LOFAR) instrument is a small scale pathfinder of SKA (Square Kilometer Array), running at lower frequencies. Its team is facing a data management challenge (producing 1 to 2 PB of reduced data per year). It is supported by the DICE-EOSC project, with the provisioning of cold storage on tapes, for a total volume of 10 PB. The host infrastructure is located in Germany (MPCDF), and is part of EUDAT/B2SAFE. The sustainability of this setup has been discussed, and a volume of 1PB/yr of new data will be stored for free, thanks to an agreement between the NenuFAR team and the MaxPlanck Institüt für RadioAstronomie.

- VAMDC (Virtual Atomic and Molecular Data Center, `https://vamdc.org`) is a long standing project with uses the EUDAT (`https://eudat.eu`) services and well as Zenodo (`https://zenodo.org`) repository for minting PIDs and storing results from their Query Store.

For most of those development, support from an EOSC-funded institution was available and really helped to level up the skills of the teams involved. We list below a set of calls available (as of fall 2023) for onboarding EOSC:

- DICE-EOSC: `https://www.dice-eosc.eu/call-service-requests`

- EGI-ACE: `https://www.egi.eu/egi-ace-open-call/`
- EOSC-Future: `https://eoscfuture-grants.eu`
- EOSC-Portal: `https://eosc-portal.eu/eosc-early-adopter-programme`
- FAIR-IMPACT: `https://fair-impact.eu/fair-impact-open-calls-support`

4 Conclusions

EOSC is being developed as a major part of Open Science in Europe. Its interactions with the astronomy field are bidirectional: astronomy projects are connecting existing services, tutorials and datasets to the EOSC so that they become part of the wider data sharing framework, and astronomy projects can benefit from the more general services offered to the whole scientific community through the EOSC. We have shown that the connection of the astronomical VO to the EOSC enables the integration of a large amount of resources to the EOSC using a system that is already well established in astronomy. These efforts represent an investment for the future where we expect EOSC will provide new capabilities for data-sharing. It is clear that engagement with these initiatives is essential so that the input and needs of the astronomical community are taken into account in the on-going shaping of the EOSC.

The ESCAPE project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement no. 824064. The EOSC Future project is co-funded by the European Union Horizon Programme under Grant Agreement no. 101017536. The Europlanet-2024 Research Infrastructure project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149. This work used the EGI Infrastructure with the dedicated support of IN2P3-IRES and CESNET- MCC. The FAIR-IMPACT project has received funding from the European Union Horizon Programme under grant agreement No 101057344.

References

Allen, M., Lamanna, G., Espinal, X., et al. 2022a, in Astronomical Society of the Pacific Conference Series, Vol. 532, Astronomical Society of the Pacific Conference Series, ed. J. E. Ruiz, F. Pierfedereci, & P. Teuben, 113

Allen, M. G., Amodeo, S., Bonnarel, F., et al. 2022b, in Bulletin of the American Astronomical Society, Vol. 54, 2022n2i028

Baumann, M. & Boch, T. 2023, cds-astro/aladin-lite: 3.2.0

Baumann, M., Pineau, F.-X., Boch, T., & Marchand, M. 2023, cds-astro/mocpy: Release v0.13.0

Cecconi, B., Philippe, H., Shih, A., et al. 2020, in European Planetary Science Congress, EPSC2020–1015

Chaudhary, S. S., Toivonen, A., Waratkar, G., et al. 2023, Low-latency gravitational wave alert products and their performance in anticipation of the fourth LIGO-Virgo-KAGRA observing run

Erard, S., Cecconi, B., Le Sidaner, P., et al. 2020, Defense Science Journal, 19, 22

Fernique, P., Nebot, A., Durand, D., et al. 2022, MOC: Multi-Order Coverage map Version 2.0, IVOA Recommendation 27 July 2022

Greco, G., Punturo, M., Allen, M., et al. 2022, Astronomy and Computing, 39, 100547

Marchand, M., Amodeo, S., & Allen, M. 2022, working paper or preprint hal-03970692

Marchand, M. C., Amodeo, S., Baumann, M., et al. 2023, Access and use of astronomy-related data from Python : a series of Jupyter notebooks tutorials Zenodo 10.5281/zenodo.7547627

Molinaro, M., Allen, M., Bonnarel, F., et al. 2021, arXiv e-prints, arXiv:2111.14468

Pineau, F.-X. & Baumann, M. 2021, MOC Lib Rust, MOCCLi, MOCWasm and MOCSet, Release Notes: https://github.com/cds-astro/cds- moc-rust/blob/main/CHANGELOG.md

Rankine, A. L., Aird, J., Ruiz, A., & Georgakakis, A. 2023

Sweijen, F., Lyu, Y., Wang, L., et al. 2023, A&A, 671, A85

Wang, Y., Murphy, T., Lenc, E., et al. 2023, Monthly Notices of the Royal Astronomical Society, 523, 5661