PROCEEDINGS OF SCIENCE

Open Science in KM3NeT

Jutta Schnabel[a,*] **for the KM3NeT collaboration**

[a] *Erlangen Centre for Astroparticle Physics, FAU Erlangen-Nürnberg, Nikolaus-Fiebiger-Str. 2, 91058 Erlangen, Germany*

E-mail: jutta.schnabel@fau.de

The KM3NeT neutrino detectors are currently under construction at two locations in the Mediterranean Sea and starting to produce valuable data both for the astrophysics and neutrino physics communities. Having committed itself to an Open Science policy, the KM3NeT collaboration is establishing a system to facilitate data sharing, open software development and integration of KM3NeT analyses in common analysis platforms, as well as providing training material. In this contribution, the current prototype architecture and future development initiatives are presented, and the benefit of Open Science for education and outreach highlighted.

XVIII International Conference on Topics in Astroparticle and Underground Physics (TAUP2023)
28.08.-01.09.2023
University of Vienna

*Speaker

PoS(TAUP2023)327

1. Open Science in the construction phase

Leaps in the development of information technology and the need to counteract the monopolisation of science communication by few publishing companies lead in the last decades to a movement that went from open access publication over the recognition of the need for open data and open software in science to the concept of "open science". Although the term is widely used, the actual definition of the goals and, even more, actual implementation of open science concepts in the day-to-day scientific practise is a challenge that is only slowly mastered step-by-step. With the UNESCO Recommendations [1] issued in 2021, the strong interlink between providing open knowledge in a dedicated open infrastructure and the societal involvement for science was highlighted, which shows the incentives to think about education and outreach and open science in a more correlated manner. It also calls on member states to foster the establishment of an open science infrastructure.

The KM3NeT collaboration has responded to the call for establishment of an Open Science policy also with the setup of a committee to implement the policy for the science practise of the future neutrino detectors. With this, the understanding of how open science impacts on daily business and which measures are needed to enable an open-science friendly setup for a large astroparticle experiment is slowly growing in the collaboration. The fields of action here can roughly be divided in the areas of providing infrastructure, see Fig. 1, and standards for the sharing of the scientific workflow, establishing quality control and procedures inside the collaboration to ensure the generation of reusable open science products, and developing policies to foster sound scientific practices.

Setting up infrastructure In order to render data, software and generally open science products FAIR[1], the data must be easy to find, retrieve and use. To increase findability, the integration of the products into large-scale common platforms or aggregators is aimed for, be it Zenodo[2] for data, Github[3] for software or Zooniverse[4] for citizen science projects.

The development of a common infrastructure in the context of the EOSC[5] helps here to find community standards and facilitates the establishment of common interfaces. The use of full KM3NeT data releases in the future might e.g. depend on the integration of the ESCAPE data lake. In the meantime, KM3NeT has established a prototype data server, the KM3NeT Open Data Centre[6], and a dedicated interface to the Virtual Observatory (VO)[7].

While these serve example data, software is developed and made available in accordance to current common practices using a dedicated KM3NeT GitLab instance and providing software containers. Software packages for reading KM3NeT data are developed. All information is made available on the KM3NeT Open Science Portal, and dedicated courses offered in the Education Portal. Here, synergies can already be seen between education and open science, as not only

[1] Findable, Accessible, Interoperable, Reusable, see https://www.go-fair.org/fair-principles/
[2] https://zenodo.org/communities/km3net
[3] https://github.com/KM3NeT/
[4] https://www.zooniverse.org/
[5] European Open Science Cloud, https://eosc-portal.eu/
[6] http://opendata.km3net.de
[7] at http://vo.km3net.de

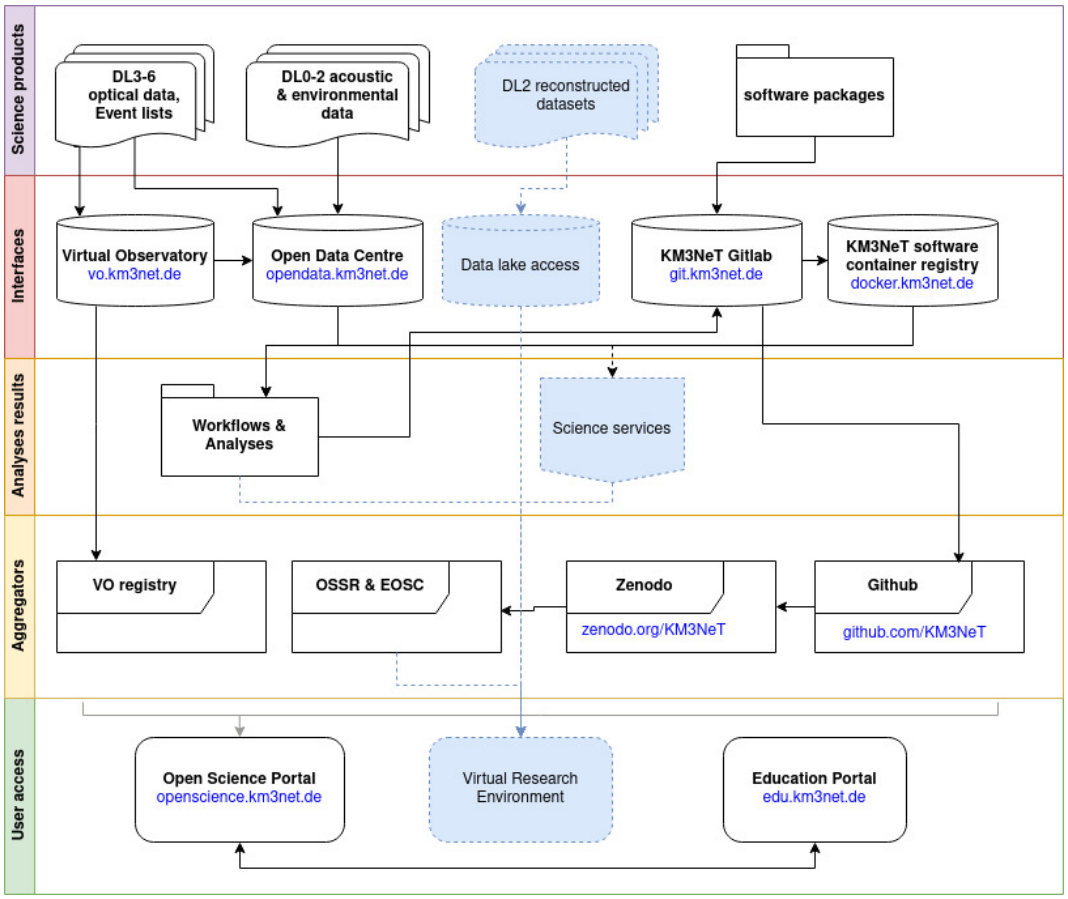

Figure 1: Overview over the KM3NeT Open Science System, starting with the science-ready data products including neutrino event detection and deep-sea environmental from lower data levels (DL). Blue components indicate possible future extensions.

internal courses from various KM3NeT boot camps are available, but also open courses based on easy-to-use Jupyter notebook examples and the online observatory of the VO.

Producing open science products As open data should not come as an additional burden to the efforts of a collaboration but as a chance well used, it is necessary to consider the publication of data as an integral part of research data management also inside the collaboration and thus include it in the Data Management Plan (DMP) [2]. In KM3NeT, the development of the data processing infrastructure, which will be based on multi-site grid computing, also already includes considerations on how publication of large-volume data sets can be facilitated and the data managed through the whole life cycle including data revisions and archiving. This especially includes quality control procedures for both data and software, and the development of public data formats into which the conversion of the internal data can easily be managed to provide coherent data releases.

Developing policies Thinking about open science in a collaboration requires also organizational changes, especially when it comes to understanding not only paper publications as outcomes of the scientific process, but making the process itself reproducible. In the end, this incentivizes

transparency of analysis projects inside the collaboration at an early stage, as well as coordination of analysis procedures and the work on common data sets developing common software.

At the publication level, review and long-term maintenance procedures have to be established for the quality control and curation of the open science products. This benefits in the end all members of the collaboration, as the technical barrier to effectively start working in the complex scientific environment is lowered through clear standardisations and improved alignment of analysis efforts. Introducing new members and students to this shared environment is even more facilitated by the fact that students that are not yet members can be trained using the open science environment and might thus be motivated to join the collaboration at a later stage of their education.

2. Open Science for Education and Outreach

The overlap between education and open science and outreach is apparent, especially when considering a step-by-step guidance of young people to enjoy engaging in science. While e.g. some visualization of open data can surely be used for outreach purposes as well, providing easy-to-follow science examples from open data can be helpful already for high school science projects, and can e.g. be offered through the collaboration education portal in KM3NeT. With increasing difficulty, this can be followed by material relevant for undergraduate projects and master classes up to graduate school environments or data challenges for the advanced science enthusiast. Using common data sets and infrastructure also intended for the use of researchers increases synergies while at the same time lowering the threshold for external scientists using open data by increasing the amount of available documentation and introduction material.

Thinking the two fields together also helps to take the addressees of outreach and education seriously by lowering the boundary between information provided for the non-experts and the one for the scientific community. On the professional level, providing open science products helps to engage in cooperation with other communities that are on the fringe of an astroparticle physicist's agenda, like open software development, big data handling and artificial intelligence, and other aspects of the advanced information age, to name only a few. Being able to show as realistically as possible what a scientist does can here be only beneficial for all those involved. On the other hand, this also fosters the idea to make open science as easily usable as possible, taking good care of the quality of example material and a thorough documentation. In the end, linking the fields already in the construction phase of a collaboration helps to decrease efforts to engage the different recipients of open research and education and outreach later.

References

[1] UNESCO, UNESCO Recommendation on Open Science, 2021 (SC-PCB-SPP/2021/OS/UROS), DOI: https://doi.org/10.54677/MNMH85460.

[2] The KM3NeT collaboration, D4.1 KM3NeT Data Management Plan, https://www.km3net.org/km3net-eu-projects/km3net-infradev/project-outputs/ (currently under revision).