KM3NeT INFRADEV – H2020 – 739560

KM3NeT report on ASTERICS liaison

KM3NeT-INFRADEV GA DELIVERABLE: D4.2

Document identifier:	**KM3NeT-INFRADEV-WP4-D4.2_final**
Date:	**15/10/2017**
Work package:	**WP4**
Lead partner:	**FAU**
Document status:	**Final version**
Dissemination level:	**Public**
Document link:	

Abstract

One of the major goals of both the ASTERICS and the KM3NeT-INFRADEV projects are to foster the dissemination of scientific results to the scientific community and citizen scientists. As part of the European Horizon 2020 initiative, both projects investigate and implement concepts for open and public data access, common standards and public software. While the ASTERICS consortium is pursuing those activities from the perspective of an initiative spanning several research infrastructures in astrophysics and astroparticle physics, KM3NeT-INFRADEV develops the means to implement common solutions on the level of an individual research infrastructure. In this document, we describe the liaison between both projects underlining specific benefits on the conceptual and executive levels, with a focus on open data access.

I. Copyright notice

II. Delivery slip

	Name	Partner/WP	Date
Author(s)	S. Geißelsöder, K. Graf and U. Katz	FAU / WP4	16/09/2017
Approved by	PMB and KM3NeT IB		03/10/2017

III. Document log

Issue	Date	Comment	Author/Partner
1.0	11/09/2017	Initial draft derived from KM3NeT-INFRADEV-WP4-D4.1_v1.4	S. Geißelsöder/FAU
1.1	16/09/2017	Updated with feedback of WP4 members	K. Graf, U. Katz/FAU
1.2	19/09/2017	Include remarks of IB referee	K. Graf/C. Bozza
1.3	21/09/2017	Include remarks of WP4 coordinator	K. Graf/U. Katz
1.4	25/09/2017	Include remarks of the project coordinator	K. Graf, U. Katz/R. van der Meer
1.5	15/10/2017	Include comments from KM3NeT Collaboration (in particular P. Coyle, E. de Wolf, K. Tsamarioudaki, K. Graf)	U. Katz

IV. Application area

This document is a deliverable for the grant agreement of the project, applicable to all partners in the KM3NeT-INFRADEV project, beneficiaries and third parties, as well as its collaborating projects.

V. Terminology

ARCA = Astroparticle Research with Cosmics in the Abyss (KM3NeT neutrino astroparticle physics telescope)
EOSC = European Open Science Cloud

IB	=	Institutional Board
IVOA	=	International Virtual Observatory Alliance
ORCA	=	Oscillation Research with Cosmics in the Abyss (KM3NeT neutrino particle physics detector)
PMB	=	Project Management Board
RI	=	Research Infrastructure
VO	=	Virtual Observatory

VI. List of figures

none

VII. List of tables

none

VIII. Project summary

KM3NeT-INFRADEV

KM3NeT is a large Research Infrastructure that will consist of a network of deep-sea neutrino telescopes in the Mediterranean Sea with user ports for Earth and Sea sciences. Following the appearance of KM3NeT 2.0 on the ESFRI roadmap 2016 and in line with the recommendations of the Assessment Expert Group in 2013, the KM3NeT-INFRADEV project addresses the Coordination and Support Actions (CSA) to prepare a legal entity and appropriate services for KM3NeT, thereby providing a sustainable solution for the operation of the research infrastructure during ten (or more) years. The KM3NeT-INFRADEV is funded by the European Commission's Horizon 2020 framework and its objectives comprise, amongst others, the preparation of Open Data Access (work package 4).

ASTERICS

The Astronomy ESFRI and Research Infrastructure Cluster, ASTERICS, is funded by the European Commission's Horizon 2020 framework. ASTERICS aims to address the cross-cutting synergies and common challenges shared by the various astronomy ESFRI facilities (SKA, CTA, KM3NeT & ELT). It brings together, for the first time, the astronomy, astrophysics and particle astrophysics communities, in addition to other related research infrastructures. ASTERICS is comprised of five work packages: ASTERICS Management Support Team (AMST), Dissemination, Engagement and Citizen Science (DECS), OBservatory E-environments Linked by common ChallengeS (OBELICS), Data Access, Discovery & Interoperability (DADI), Connecting Locations of ESFRI Observatories and Partners in Astronomy for Timing and Real-time Alerts (CLEOPATRA).

IX. Executive summary

One of the major goals of both the ASTERICS and the KM3NeT-INFRADEV projects are to foster the dissemination of scientific results to the scientific community and citizen scientists. As part of the European Horizon 2020 initiative, both projects investigate and implement concepts for open and public data access, common standards and public software. While the ASTERICS consortium is pursuing those activities from the perspective of an initiative spanning several research infrastructures in astrophysics and astroparticle physics, KM3NeT-INFRADEV develops the means to implement common solutions on the level of an individual research infrastructure. In this, valuable experience is shared between KM3NeT and its predecessor project ANTARES, which is also represented in ASTERICS. In this document, we describe the liaison between KM3NeT-INFRADEV and ASTERICS concerning synergetic cooperation on the conceptual and executive levels, with a focus on open data access. In particular, we identify objectives that are pursued by both projects, goals that can be reached in a common, concerted effort, and an executive scenario in which these plans are to be implemented.

Table of Contents

1 Introduction

KM3NeT is a large Research Infrastructure (RI) that will consist of a network of deep-sea neutrino detectors in the Mediterranean Sea with user ports for Earth and Sea sciences. The main science objectives, a description of the technology and a summary of the costs are presented in the KM3NeT 2.0 Letter of Intent (1).

KM3NeT will open a new window on our Universe, but also forward the research into the properties of neutrinos. With the ARCA telescope at the KM3NeT-It installation site near Sicily, KM3NeT scientists will search for neutrinos from distant astrophysical sources such as supernovae, gamma ray bursts or active galactic nuclei. Using the same technology, the ORCA detector at the KM3NeT-Fr installation site near Toulon will provide data of unprecedented quality on neutrino oscillations, exploiting neutrinos generated in the Earth's atmosphere. Arrays of thousands of optical sensors will detect the faint light in the deep sea from charged particles originating from collisions of the neutrinos with atomic nuclei. At both installation sites, the facility will also house instrumentation for Earth and Sea sciences for long-term and on-line monitoring of the deep-sea environment and the sea bottom at depth of several kilometres (2).

The KM3NeT Collaboration has developed a data policy plan (3) reflecting the research, educational and outreach goals of the facility. For a certain embargo time (e.g. two years) after data taking, the processing and exploitation of the data is granted to the Collaboration members as a return for constructing, maintaining and operating the facility. During this phase, each Collaboration member has full access rights to all data, software and know-how. The Collaboration commits itself to process the data during the embargo phase so as to generate high-quality calibrated and reconstructed event data suited for a wider user community. These data will be made publicly available after the embargo time under an open-access policy on a web-based service and will not only allow the scientific community to validate the results presented by the Collaboration but also to perform individual analyses.

The prompt dissemination of scientific or methodological results achieved during the embargo time is in the responsibility of the KM3NeT Collaboration. The scientific responsibility and the publication rights for results derived from public data is with the scientists performing the corresponding analyses. Subject to the availability of resources, the KM3NeT Collaboration offers analysis support to external analysers on their request, and after scrutinising the validity of the respective analyses. In this case, both the external scientists and the KM3NeT Collaboration will author the resulting publications.

This document details the established connections between ASTERICS and KM3NeT including examples where this liaison creates and exploits synergies between the two projects. It contains information on specific opportunities, the approaches suggested or already under development, and how these aid the implementation of open data, its preservation and curation.

One of the aims of this project is to provide a sustainable basis for the development and utilisation of e-infrastructure commons. KM3NeT will therefore contribute to the development of standards and services in the e-infrastructures both in the specific research field and in general. The guidelines for

these developments are recommendations by expert groups (ASPERA, e-IRG, RDA, W3 or similar) and a best-practice approach.

2 Existing links between ASTERICS and KM3NeT-INFRADEV

The aspired liaison is facilitated by strong links between both projects on the organisational and management levels, and by institutes and personnel that are active in both projects.

2.1 Institutions

Several of the major institutions in the KM3NeT Collaboration have joined the ASTERICS consortium. Out of the six project partners of KM3NeT-INFRADEV, four are partners in ASTERICS as well, namely the Netherlands Research Organisation (NWO) represented by Nikhef, Amsterdam, The Netherlands; the Centre National de la Recherche Scientifique (CNRS) represented by the Center for Particle Physics of Marseille (CPPM), Marseille, France; the German Friedrich-Alexander University Erlangen-Nürnberg (FAU) represented by the Erlangen Centre for Astroparticle Physics (ECAP), Erlangen, Germany; the Italian Istituto Nazionale di Fisica Nucleare (INFN) represented by the Laboratori Nazionali del Sud (LNS), Catania, Italy.

2.2 Personnel

Scientists of the institutions listed above are strongly committed in both projects, ensuring a continuous cooperation. The individuals involved include Maarten de Jong (NWO/Nikhef) as KM3NeT-INFRADEV project coordinator and Uli Katz (FAU/ECAP) as WP4 work package leader. In addition, both projects share the project coordinator Rob van der Meer (ASTRON/Nikhef).

Close cooperation is also established in day-to-day work. In particular at FAU/ECAP, the coordinating institute of KM3NeT-INFRADEV WP4, personnel from both projects work side-by-side in the same offices and cover tasks relevant to both projects.

The close links are also demonstrated by repeated invitations of KM3NeT colleagues to ASTERICS activities, e.g. a presentation of KM3NeT computing at the first ASTERICS-OBELICS Workshop, December 2016, in Rome, Italy or lectures by KM3NeT experts at the first ASTERICS-OBELICS International School, June 2017, in Annecy-le-Vieux, France. In other fields, such as the VO technical forums, the cooperation will be ramped up in the context of the KM3NeT-INFRADEV project.

3 Synergy options offered by cooperation between ASTERICS and KM3NeT-INFRADEV

In this chapter, we discuss concrete, expected or already established cases where KM3NeT-INFRADEV is utilising concepts investigated within ASTERICS or where specific developments within KM3NeT will serve as important input to the success of the ASTERICS project.

3.1 Compatibility with the Virtual Observatory

The Virtual Observatory is an international astronomical community-based initiative. It aims to establish global electronic access to available data archives of space and ground-based astronomical observatories and other sky survey databases. EURO-VO targets the deployment of an operational VO in Europe. It supports the use of VO tools and services by the scientific community, technology take-up and VO compliant resource provision, and building the technical infrastructure (4).

The VO is publicly accessible for anyone interested, professional or citizen scientist. Due to its flexible design, it allows for employment of the innovative multi-messenger approach that promises new insights for various kinds of observations, especially – but not limited to – the field of astronomy, astrophysics and astroparticle physics represented in the ASTERICS project.

Using the guidelines developed in ASTERICS, KM3NeT-INFRADEV will establish the means to convert public data generated by KM3NeT to file formats suited for the VO and to guarantee that the released data will be fully accessible in the VO.

3.2 Lightweight containerisation and data persistence

The processing and analyses of data from research infrastructures need to be reproducible for a long time, i.e. both data and analysis software have to be preserved over a time span of typically longer than 10 years. An idea developed in the context of ASTERICS is to use lightweight containerisation[1] for this task. Adopted to open data access, this would allow not only for the release of public data as an end-product, but also for publishing processes that generate these data and the methods used to derive results from these data. This approach facilitates the future reproducibility of published analyses and allows for cross-checks, easier transfers of analysis methods, and easier updates.

Compared to other virtualisation techniques like virtual machines, a key advantage is less overhead, especially in terms of storage size. Docker (5) and Singularity (6) have been identified as suited technologies that can be used for this purpose. The knowledge on how to utilise these technologies for this purpose has been studied within ASTERICS and the applicability to KM3NeT open access data will be studied within KM3NeT-INFRADEV.

[1] Containerisation is an alternative to full machine virtualisation that involves encapsulating an application in a software container with its own operating environment.

3.3 Common file formats

To facilitate easy exchange of public data or even interoperability of methods between RIs, the ability to read and write common data formats is essential. Both the HDF5 file format (7) and formats based on FITS (8) have been supported within ASTERICS and are therefore considered reasonable candidates for the target file format for public data within KM3NeT-INFRADEV. Open data access needs a high level of data quality to ensure consistency. At the same time, interoperability and flexibility are a necessary must. To ensure this, we intend to publish the data in one selected efficient format (HDF5 or FITS) and to provide thoroughly tested conversion mechanisms. This approach allows us to support more file formats, e.g. the simple CSV file format for smaller data sets, the common ROOT (9) data format or special file format solutions for the VO. At the same time, data duplication, and potential inconsistencies between data sets produced in different formats at different stages of a processing chain can be avoided, while keeping the efficiency of an optimised data format.

KM3NeT-INFRADEV will serve as a testbed for the implementation of this strategy along the general guidelines and recommendations provided by ASTERICS.

3.4 European Open Science Cloud

Modern astrophysics and astroparticle physics RIs have increasing computing demands in terms of data storage and computing power. At the same time, they must cope with the fact that the available computing resources are distributed and inhomogeneous in their set-up. A unified, cloud-based approach provides the most efficient way to utilise the resources; it helps to process data reliably and reduces the potential for errors and delays. With the envisioned European Open Science Cloud (EOSC), it will become simpler for scientists to not only use the computing power efficiently, but also to transfer and adapt a workflow established in one RI to another. The possibilities of the EOSC as well as the hurdles on the way to such a service need to be investigated. Some ASTERICS partners participate in the EOSC pilot phase and will share the experience within the ASTERICS consortium. KM3NeT-INFRADEV as well as all ASTERICS partner RIs will profit from a common approach towards the EOSC, for example concerning the storage of public data, but also for standards in the conducted analyses.

3.5 Shared expertise for licensing of open data

Publishing open data has not been as widespread a practice within the astroparticle physics community as it has already been for many years in astronomy. Besides the obvious technical solutions that need to be established for this step, the issue of licensing the published data is a relevant aspect where KM3NeT-INFRADEV can learn from the expertise and experiences of other large collaborations. We will therefore evaluate the licensing models already present in other major RIs involved in ASTERICS to identify the best suited model to use for public neutrino astronomy data.

A pathfinding effort was already done by the CPPM group within the ASTERICS-DADI project, where data of the neutrino telescope ANTARES was uploaded to the VO by GAVO (10). This know-how will directly be used for the publication of open access data with KM3NeT-INFRADEV.

3.6 Development of universally applicable software

Software developed within one collaboration, but with the requirements of other RIs in mind, can be designed more flexibly to be of general value for the scientific community. The extended cooperation between KM3NeT and other groups has already fostered such examples.

3.6.1 CORELib

The Library of cosmic ray events (CORELib) developed in ASTERICS deals with the recurring need for a high number of simulated interactions of cosmic rays with the atmosphere of the Earth. While many RIs are confronted with this problem and invest significant amounts of computing resources for this task, there is no benefit from repeating this process uncoordinatedly and therefore inefficiently. A well-tested, centralized repository providing these simulations does not only save time and costs in terms of computing power, but also prevents non-obvious errors resulting from mistakes or inconsistencies when using the simulation software. CORELib uses the established CORSIKA framework and documents all parameters that go into the simulations. In terms of open data this approach fosters the reproducibility of studies, as a common, well-referenceable input of simulations and allows this part of an analysis to be reproduced by other groups. In KM3NeT, CORELib is already in intense use and the experience is reported back to ASTERICS on a regular basis.

3.6.2 ROAst

The ROOT extension for Astronomy and astrophysics (ROAst) is a package written in ROOT to add capabilities and tools for astrophysical research. The main contributions are interfaces to import and manipulate astronomical catalogues and to create a set of interfaces as bridge to other software such as data-sets and numerical simulation software for relevant astrophysical phenomenology. It has been developed within the ASTERICS project with KM3NeT as test case RI, but with the aim to be a valuable tool for most RIs making use of catalogues of astronomical objects or astronomical coordinate systems in general. In KM3NeT-INFRADEV WP4, it is foreseen to design an interface between the open access data and ROAst.

3.6.3 JPP

Another pathfinding project within ASTERICS, performed by the CPPM group, is making the main data processing and analysis software package - JPP - available to the public. The license and publication rules adopted for this project will be basis for making available all the software to analyse the public data of KM3NeT-INFRADEV

4 Cooperation between ASTERICS and KM3NeT-INFRADEV

Exploiting the synergies discussed above requires a continuous and close cooperation, which is already established and will be strengthened, building on the institutional and personnel links presented in section 2. As a general guideline, KM3NeT-specific work (i.e. tasks from which mostly

KM3NeT will benefit) is to be pursued in the KM3NeT-INFRADEV framework, whereas general developments targeting several RIs or the whole field of astrophysics are performed in ASTERICS. It is, however, clear that assignment of tasks to the one or other project is not unambiguous and needs to be discussed and decided on a case-by-case basis.

To facilitate such organisation, both projects have agreed on a formalised cross-coordination at the management level, where the liaison and sharing of work will be permanently on the agenda. This will be implemented and actively pursued by the common project coordinator and will entail a coordinated cooperation at the operational level. One example where such a process is already established is the use and further development of CORELib (see section 3.6.1): Inside KM3NeT this work is pursued and organised by scientists who are also part of the corresponding ASTERICS work package and regularly report back to ASTERICS. Similar organisational structures are in place (ROAst, see section 3.6.2) or will be established.

5 References

1. *Letter of Intent for KM3NeT 2.0.* **Adrián-Martínez, S., et al.** 2016, Journal of Physics G: Nuclear and Particle Physics, Vol. 43 (8), p. 084001. arXiv:1601.07459 [astro-ph.IM]. DOI: 10.1088/0954-3899/43/8/084001.

2. **The KM3NeT Collaboration.** KM3NeT Homepage. [Online] 2017. [Cited: 29th May 2017.] https://www.km3net.org.

3. **KM3NeT-InfraDev.** *The KM3NeT Data Management Plan.* KM3NeT-InfraDev. 2017. KM3NeT2.0_WP4_D4.1.

4. **The European Virtual Observatory EURO-VO.** EURO-VO webpage. [Online] http://www.euro-vo.org.

5. **Docker.** The Docker webpage. [Online] https://www.docker.com.

6. **Singularity.** The Singularity webpage. [Online] http://singularity.lbl.gov.

7. **The HDF Group.** HDF5 support page. [Online] https://support.hdfgroup.org/HDF5/.

8. *FITS: A Flexible Image Transport System.* **Wells, D. C., Greisen, E. W. and Harten, R. H.** 06 1981, Astronomy and Astrophysics Supplement Series, Vol. 44, pp. 363-370.

9. **CERN.** ROOT Data Analysis Framework homepage. [Online] https://root.cern.ch.

10. **German Astrophysical Virtual Observatory (GAVO).** GAVO Data Center. [Online] http://dc.zah.uni-heidelberg.de.