KM3NeT INFRADEV – H2020 – 739560

Report on implementation and user tests of virtual education centre and on training workshops

KM3NeT-INFRADEV GA DELIVERABLE: D4.11

Document identifier:	**KM3NeT-INFRADEV-WP4-D4.11**
Date:	**18/09/2020**
Work package:	**WP4**
Lead partner:	**FAU**
Document status:	**Final**
Dissemination level:	**Public**
Document link:	

Abstract

The KM3NeT Research Infrastructure will, over a period of at least a decade, produce a large amount of unique scientific data that are to be made available to the scientific communities concerned and to the broader general public. This requires the set-up of tools, procedures, documentation and rules to provide this service. Training on how to handle and analyse the data is essential for external usage. This training is provided through the KM3NeT Virtual Education Centre. The implementation of the virtual education centre as well as a reference to the tutorials prepared for the training workshops are described in this document.

I. Copyright notice

II. Delivery slip

	Names	Partner/WP	Date
Author(s)	Dimitris Stavropoulos	NCSR-D/WP4	31/07/2020
Approved by	KM3NeT IB		18/09/2020

III. Document log

Issue	Date	Comment	Author/Partner
0.1	30/07/2020	First draft	D. Stavropoulos/NCSR-D
1.0	31/07/2020	Reviewed by C. Markou	D. Stavropoulos/NCSR-D
1.1	14/08/2020	Comments and edits by K. Graf & J. Schnabel	D. Stavropoulos/NCSR-D
1.2	15/09/2020	Reviewed by R. Shanidze	D. Stavropoulos/NCSR-D

IV. Application area

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

V. Terminology

ADQL = Astronomical Data Query Language
ANTARES = Astronomy with a Neutrino Telescope and Abyss environmental RESearch
ARCA = Astroparticle Research with Cosmics in the Abyss (KM3NeT neutrino astroparticle physics telescope)
CSA = Coordination and Support Actions
DaCHS = Data Center Helper Suite
ESFRI = European Strategy Forum on Research Infrastructures

FAIR	= Findable Accessible Interoperable Reproducible
GRB	= Gamma Ray Bursts
IVOA	= International Virtual Observatory Alliance
OBS	= Open Broadcaster Software
ORCA	= Oscillation Research with Cosmics in the Abyss (KM3NeT neutrino particle physics detector)
SAMP	= Simple Application Messaging Protocol
SCS	= Simple Cone Search
TAP	= Table Access Protocol
TOPCAT	= Tool for Operations on Catalogues And Tables
VO	= Virtual Observatory

VI. List of figures

VII. List of tables

None

VIII. Project summary

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).

IX. Executive summary

The KM3NeT Research Infrastructure will, over a period of at least a decade, produce a large amount of unique scientific data that are to be made available to the scientific communities concerned and to the broader general public. This requires the set-up of tools, procedures, documentation and rules to provide this service. Training on how to handle and analyse the data is essential for external usage. This training is provided through the KM3NeT Virtual Education Centre. The implementation of the virtual

education centre as well as a reference to the tutorials prepared for the training workshops are described in this document.

X. 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, KM3NeT scientists will search for neutrinos from distant astrophysical sources such as supernovae, gamma ray bursts or active galactic nuclei. Using the exact same technology, the ORCA detector 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 Cherenkov light generated in the deep sea by charged particles originating from collisions of the neutrinos with atomic nuclei. 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 a 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. During a certain embargo time (e.g. two years, to be ratified by the KM3NeT Collaboration) access to the data will be restricted to the KM3NeT Collaboration for processing and calibrating the raw data, and securing their quality and correctness. During this period, the exploitation of the data is exclusively granted to the collaboration members as a return for constructing, maintaining and operating the facility. The collaboration commits itself to generating high-quality reconstructed event data suited for a wider user community during the embargo period. These data will subsequently be made publicly available under an open-access policy on a web-based service and will not only allow the public to validate the scientific results presented by the collaboration but also to perform individual analyses.

The contribution of KM3NeT to the body of scientific knowledge will depend to a large extent on the quality of the analysed data. A data management plan that ensures a correct handling of the KM3NeT data along all the production and processing chain has been presented in [3]. From an open-access perspective, the data management plan should also be compliant with the FAIR (Findable - Accessible - Interoperable - Reproducible) data principles, which were specifically designed to enable and enhance the reuse of scholarly data by third parties [4], [5].

This document describes the implementation of the KM3NeT Virtual Education Centre furthering the prototype installation [6] and to the educational material this contains. The data provided to the broader scientific community have to be accompanied with the relevant educational material. The Virtual Education Centre has been improved to make those host materials better findable and reusable to external users. Tutorials on how to retrieve the data, on making simple analyses, as well as use cases combining the provided data with other publicly available scientific datasets have been created and are described in this document.

2. The Virtual Education Centre

The educational material of the KM3NeT collaboration is provided through the Virtual Education Centre [7] (*Figure 1*). While the first prototype [6] was based on an indico service, this was found no sufficient for the aims of a virtual education centre during user tests. There was, e.g., no video streaming service and no feedback mechanism available. The new Virtual Education Centre consists of a server running the Wordpress content management system [8]. The LifterLMS plugin [9] has been installed in Wordpress, providing a specialized education management system.

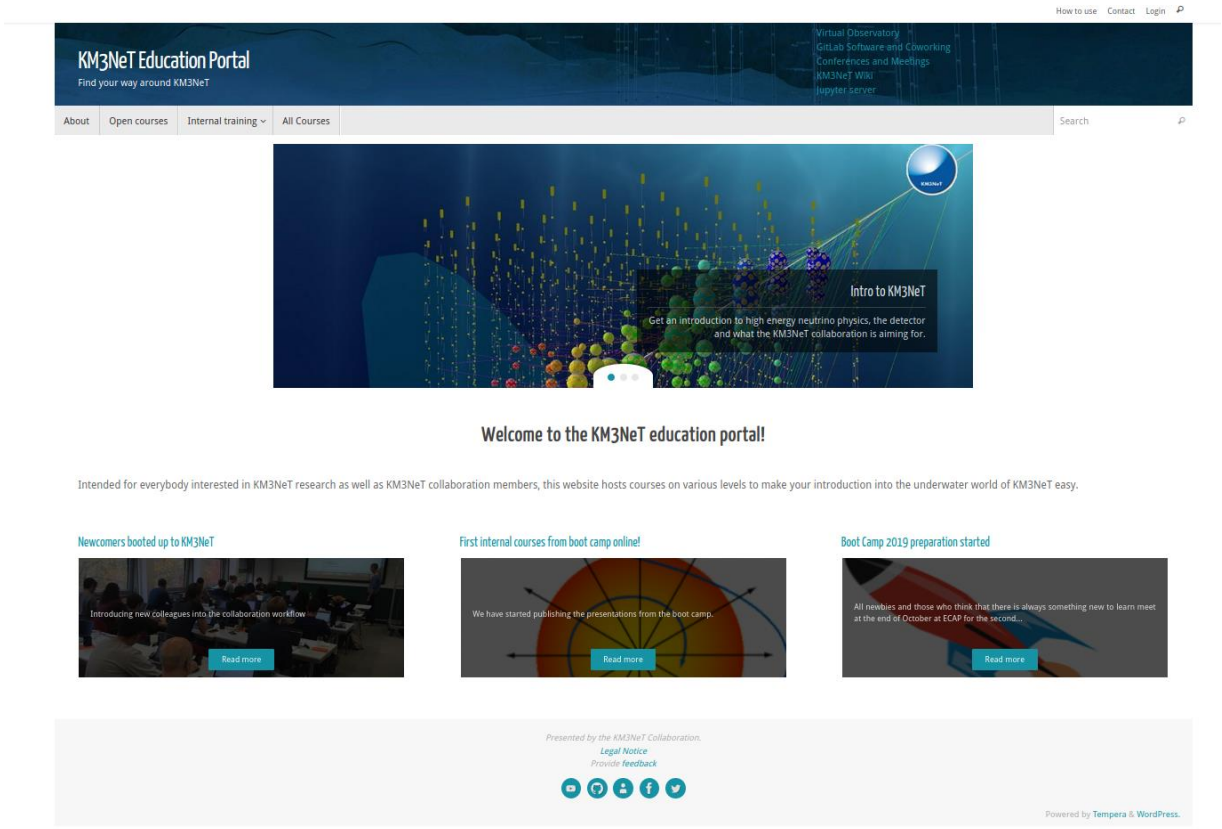

Figure 1: The home page of the KM3NeT Virtual Education Centre [6].

The courses in the Virtual Education Centre are divided in two categories. They are either protected or freely available. The courses belonging to the first category are accessible exclusively by KM3NeT members, who are able to enrol by authenticating with their KM3NeT user identity, later on the link to federated authentication and authorisation services from the EOSC are foreseen. These courses are oriented mainly to newcomers making their first steps in the collaboration. The courses of the second category are openly and freely accessible by all users, containing information on how to use the open

data provided by the collaboration and aimed to provide an introduction to scientific research with the KM3NeT detectors.

The education management system provided by the LifterLMS plugin, offers the possibility to set up the sequencing of instructions in hierarchical categories like courses, sections and lessons (*Figure 2*). The sequence in which the courses have been set up is, in principle, a reference program for face-to-face training meetings. In addition, it allows the users to follow the sequencing of instructions in a self-paced study program independent of dedicated meetings.

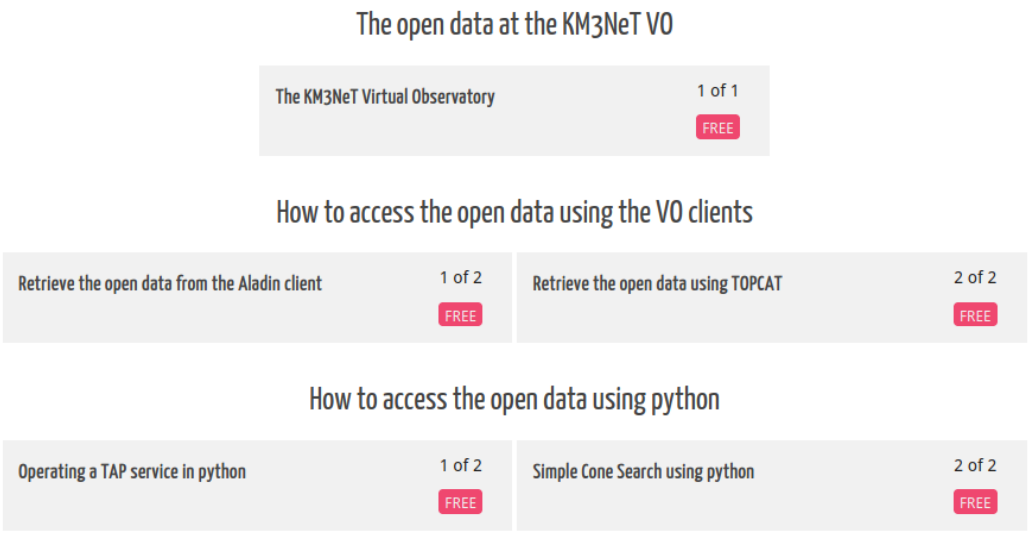

Figure 2: Exemplary lessons of an open course categorized in different sections.

3. Tutorials on using neutrino data

3.1. Neutrino events in the KM3NeT Virtual Observatory

The astrophysical open data provided by the KM3NeT collaboration are found at the KM3NeT Virtual Observatory [10]. This is handled by a dedicated server running the DaCHS software [11]. This software is designed for setting up and running a data center, compliant with the standard protocols for data exchange set by the International Virtual Observatory Alliance (IVOA) [12].

The data consist of neutrino events. An event is the set of information recorded by the infrastructure in a certain time interval, when the criteria that have been set for an event of interest, e.g. the detection of a neutrino interaction near the detectors, are satisfied. After the data processing, the aggregated event data consists of parameters such as the reconstructed direction and the detection time. The values of these parameters are tabulated. As the ARCA and ORCA detectors are currently under construction, as first test data, a dataset containing neutrino events with astrophysical origin from the ANTARES detector [13] is provided through the KM3NeT Virtual Observatory.

3.2. Protocols

Among the protocols have been set by IVOA, the Table Access Protocol (TAP) [14] defines the rules by which astronomical catalogues as well as general database tables are being accessed. The TAP allows SQL-like commands to be executed against the tables in order to access the data. Hence, commands in Astronomical Data Query Language (ADQL) [15] can be executed, taking advantage of its functions which are oriented to queries against astronomical data tables.

Another way to access datasets that comply with the protocols set by IVOA, is the Simple Cone Search (SCS) [16]. This defines a simple query protocol for retrieving records from a catalogue of astronomical sources. The query describes a central sky position and an angular distance, defining a cone on the sky. The astronomical sources whose positions lie within the cone are being retrieved.

There is a variety of graphical user interfaces that make possible to edit and visualize tabular data as well as combining data from different sources. This is possible due to the compliance of the public datasets provided from several experiments with the standard IVOA protocols. There are few so-called “TAP clients”, with the most widely used to be the Aladin [17] and the TOPCAT [18]. These TAP clients interoperate and communicate under the Simple Application Messaging Protocol (SAMP) [19]. Hence, the user is able to perform analyses taking advantage of the different options each client offers by transferring the datasets between TAP clients.

There is also an option to access the data using the python programming language. The relevant code libraries are included in the pyVO package [20]. This allows finding and retrieving astronomical data available from archives that support standard IVOA protocols.

The implementation of the KM3NeT open data system and the involved access protocols that have been demonstrated here will be fully described in deliverable D4.8.

3.3. Tutorials

The tutorials developed for the freely available courses are in the form of videos. These videos have been recorded using the Open Broadcaster Software (OBS) [21]. Each tutorial is accompanied with a hands-on exercise for the user (*Figure 3*). These exercises are similar to the content of the respective video being shown in the lesson. An introductory course is also available with the relevant information on the KM3NeT detectors.

The tutorials have been created with the aim to show to the user how to retrieve the data, how to analyse and visualise them, and finally how to combine them with open data from other sources demonstrating two different use cases. The different goals are approached using the TAP clients as well as python scripting, trying to accommodate for the different programming skills and backgrounds of users.

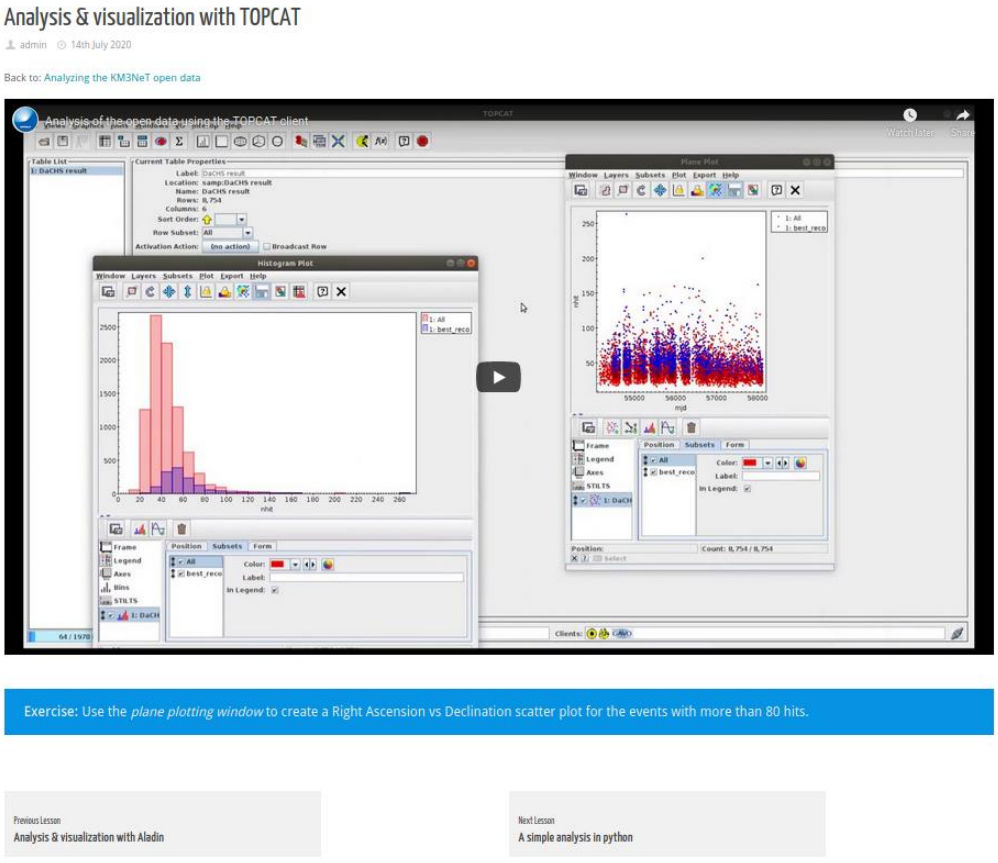

Figure 3: A tutorial video followed by the respective exercise. The tabs at the bottom allow the user to move to the previous/next lesson, following the flow of the courses.

3.3.1. Retrieving the data

Tutorials on how to retrieve the data have been developed as open courses. The open tutorials introduce the user to how to query the data tables from the KM3NeT Virtual Observatory using the ADQL and SCS forms, and how to retrieve the data using the TAP clients as well as python scripting commands. Except for the variety of options the user has in retrieving the data, these tutorials demonstrate the IVOA protocols being followed by the KM3NeT open data system.

3.3.2. Simple analyses and use cases

Several tutorials on how to handle, analyse and visualise the data have been created. These are demonstrated by using either the TAP clients, or following an example python script for users more familiar in programming. These tutorials consist of creating distributions of the parameters and performing correlations tests between them, as well as creating subsets from the initial dataset.

Having been introduced on how to analyse the data, the user experiences two use cases. The first one combines the astrophysical neutrino events with observations from Gamma Ray Bursts (GRBs), taking into account that the neutrino emission from the GRBs is a phenomenon with high scientific interest. This analysis is being performed using the TAP clients. Concerning the other uses case, a correlation with a dataset that contains neutrino candidates published by the IceCube collaboration is performed. Python commands are used for this analysis.

4. User tests and training workshops

The open courses are being tested by external scientists, mostly PhD students from other laboratories in NCSR-Demokritos, as well as, intern students. User tests from internal KM3NeT members are also important as they already possess the knowledge about the content to be introduced.

Training workshops will be organised in the near future, both face-to-face and remotely. The feedback due to the interaction with the participants will be used on improving the open courses.

5. References

1: Adrián-Martínez, S. e. (2016). Letter of Intent for KM3NeT 2.0. Journal of Physics G: Nuclear and Particle Physics, 43 (8), 084001.

2: The KM3NeT Project: https://www.km3net.org/

3: KM3NeT-InfraDev. (2017). The KM3NeT Data Management Plan.

4: The FAIR principles: Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; et al. https://www.nature.com/articles/sdata201618. *Scientific Data*. p. 160018.

5: The FAIR principles: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

6: The prototype Virtual Education Centre: https://www.km3net.org/wp-content/uploads/2020/02/D4.4-KM3NeT-virtual-education-centre.pdf

7: The KM3NeT Virtual Education Centre: https://edu.km3net.de/

8: Wordpress: https://wordpress.org/

9: LifterLMS: https://lifterlms.com/

10: KM3NeT Virtual Observatory: http://vo.km3net.de/

11: DaCHS software: http://soft.g-vo.org/dachs

12: IVOA protocols: http://ivoa.net/deployers/intro_to_vo_concepts.html

13: Neutrino dataset from the ANTARES detector: https://antares.in2p3.fr/publicdata2017.html

14: Table Access Protocol: http://www.ivoa.net/documents/TAP/

15: ADQL: http://www.ivoa.net/documents/ADQL/20180112/PR-ADQL-2.1-20180112.html

16: Simple Cone Search: http://www.ivoa.net/documents/latest/ConeSearch.html

17: Aladin: https://aladin.u-strasbg.fr/

18: TOPCAT: http://www.star.bris.ac.uk/~mbt/topcat/

19: SAMP: http://www.ivoa.net/documents/SAMP/

20: pyVO: https://pyvo.readthedocs.io/en/latest/

21: OBS: https://obsproject.com/