

# Preserving ANTARES legacy data

Jutta Schnabel and Thomas Eberl, FAU Erlangen (ECAP)

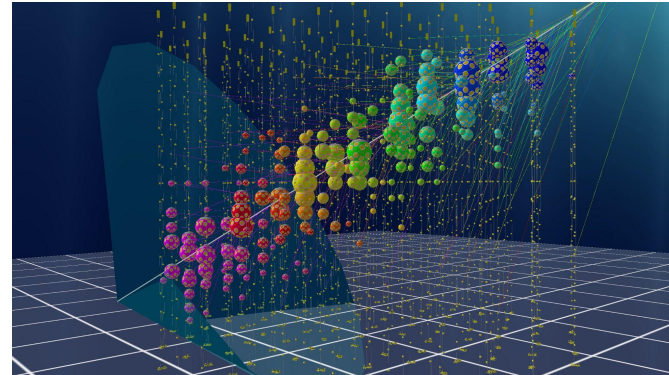
4th DPHEP Collaboration Workshop

CERN, 2<sup>nd</sup> - 3<sup>rd</sup> October 2024

# Overview

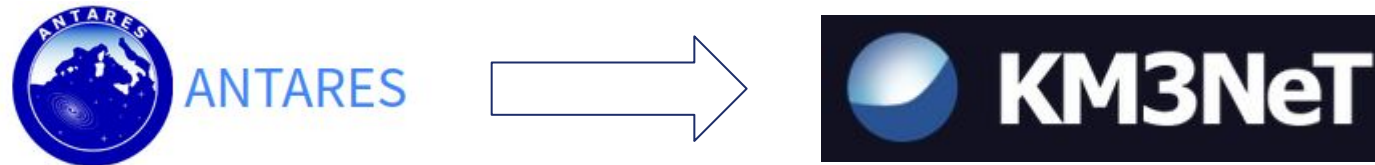
Preserving ANTARES data by making it KM3NeT data

- What is the data?

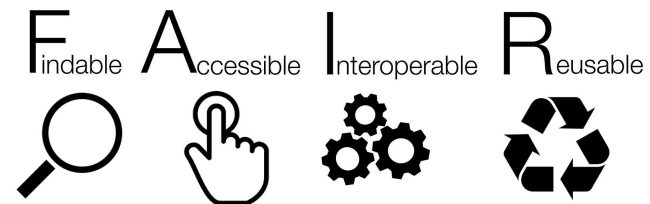


Cherenkov detection of high-energy  $\nu$

- How is it done?



- What do we get out of it?



# The ANTARES and KM3NeT detectors

High energy neutrino detection with Cherenkov detectors in the Mediterranean

## High-energy neutrino water Cherenkov detectors

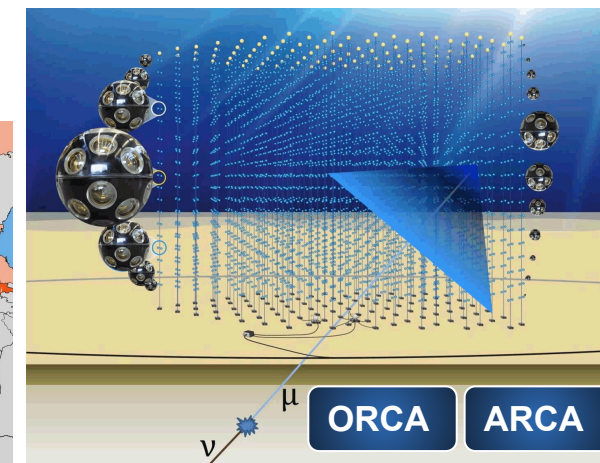
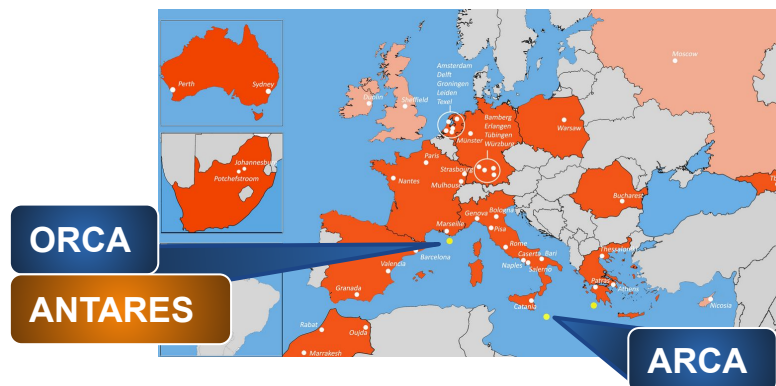
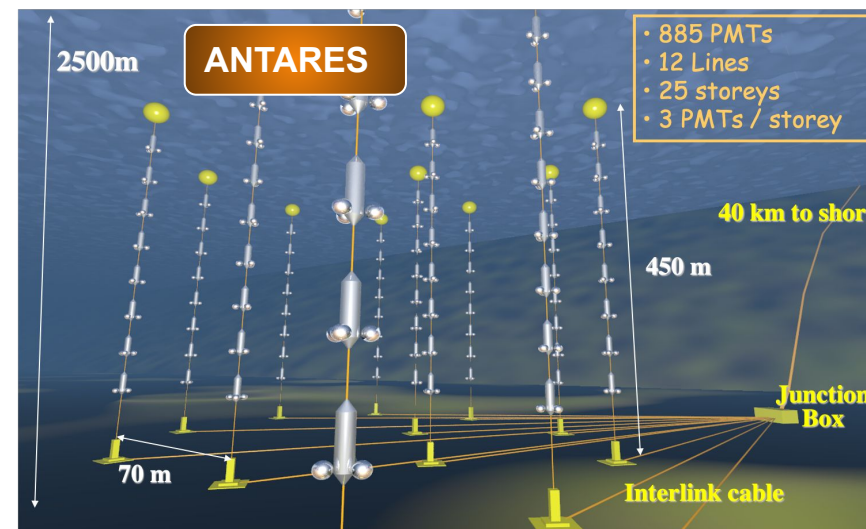
- situated in the Mediterranean Sea
- detecting single secondary photons of relativistic particles with (Digital) Optical Modules (DOMs)
- Low-countrate experiment with large atmospheric background
- Secondary systems for environmental measurements and detector calibration
- aimed at astroparticle physics and studies of neutrino properties

## ANTARES

- operational from mid-2000s until 2022 (dismantling)
- 12 detection lines covering 0.03 km<sup>2</sup>

## KM3NeT

- ORCA (neutrino properties) and ARCA (astrophysics)
- under construction, partially operational
- full configuration 3 x 115 detection lines 0(km<sup>3</sup>)

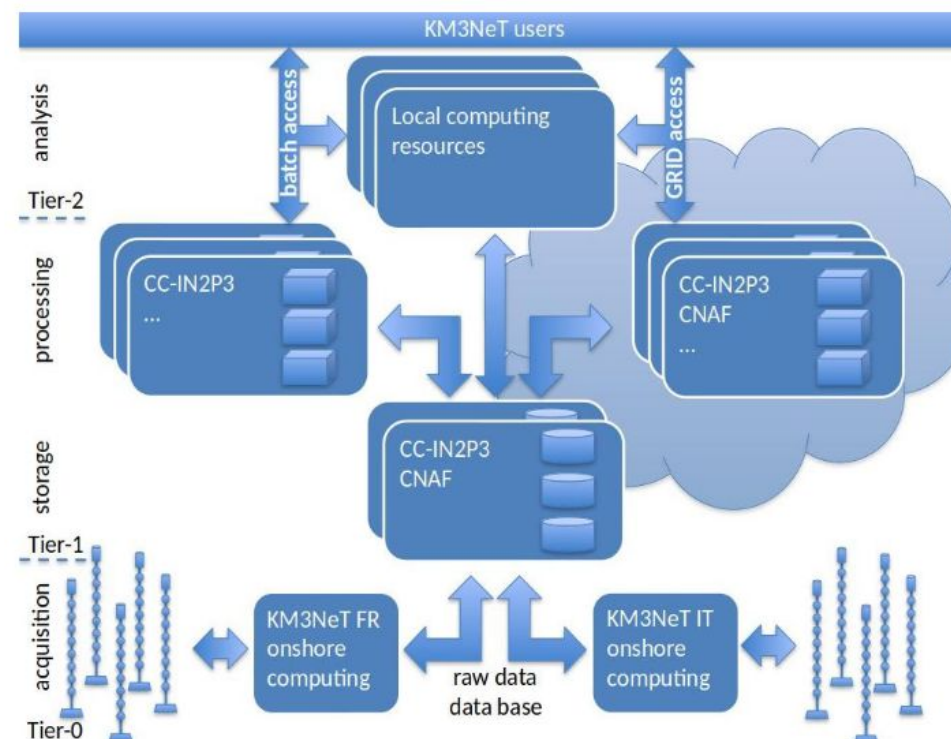
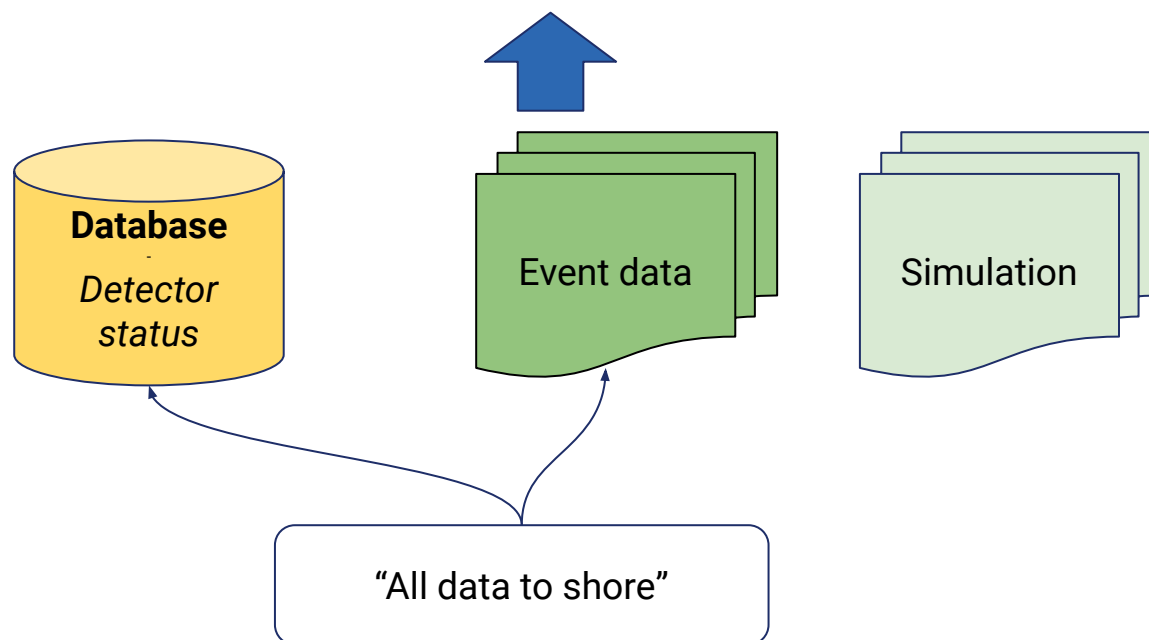




# More than neutrino events

Events, simulation and detector data

Decl [deg]	RA [deg]	Nhit [deg]	Beta	MJD [days]
19.5	68.2	21	1.0	54138.3105
-60.0	26.5	33	0.8	54138.5830
-29.8	82.1	34	0.3	54140.2299
-8.6	271.8	41	0.3	54140.6394
-32.3	261.4	45	0.5	54142.7042
-66.7	149.9	52	0.8	54159.4158



# Current status of ANTARES open data

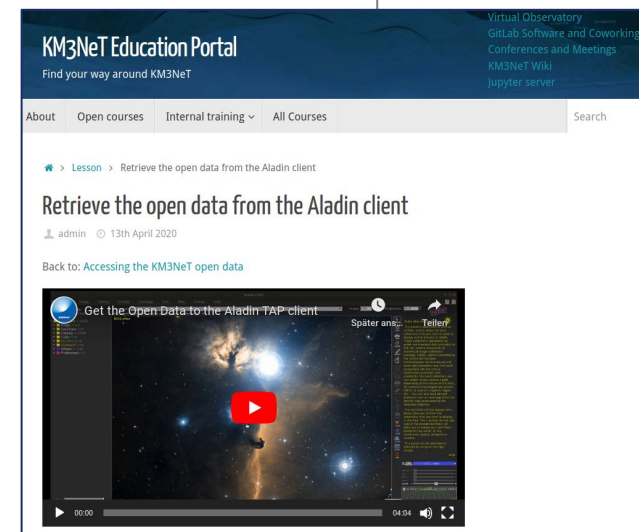
- Few data sets so far provided on webpage
- Containing event lists of  $O(10^4)$  events
- Integrated in the “Virtual Observatory”
- Plots for background estimate etc. provided
- Already integrated in KM3NeT as example data

## Data set for the 2007-2017 ANTARES search for cosmic neutrino point sources

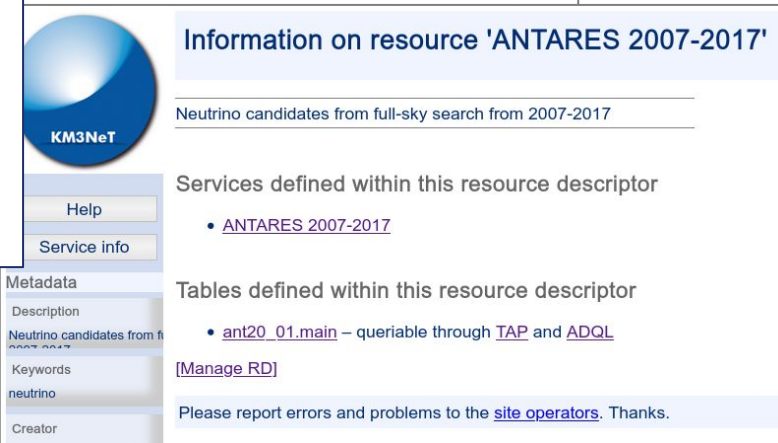
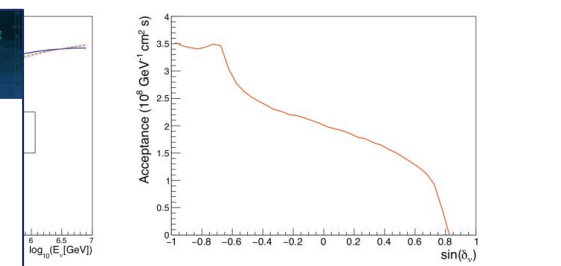
### Introduction

The present data set corresponds to the track sample (muon neutrino candidates) of a study meant to search for a point sources with data collected from January 2007 to December 2017 by the ANTARES neutrino telescope. Attached below are the effective area, the acceptance, the cumulative angular resolution distribution and the point spread function for an E-2 source spectrum. This sample encompasses the two previous released sets, with similar cuts. This increased lifetime amounts to 3125 days and to a total number of 8754 events.

More information on how the search was performed can be found in:  
[G. Illuminati for the ANTARES Collaboration, PoS\(ICRC2019\)920](#)  
[A. Albert et al., ApJL 863, L30 \(2018\)](#)



The screenshot shows the KM3NeT Education Portal interface. The main content area displays a lesson titled "Retrieve the open data from the Aladin client" by admin, dated 13th April 2020. Below the lesson title is a video player with the title "Get the Open Data to the Aladin TAP client" and a play button. The video player shows a dark sky with a bright star and a red play button in the center.



The screenshot shows the "Information on resource 'ANTARES 2007-2017'" page. The page includes a KM3NeT logo, a description of the resource as "Neutrino candidates from full-sky search from 2007-2017", and a list of services defined within the resource descriptor, including "ANTARES 2007-2017". The page also includes a metadata section with fields for Description, Keywords, and Creator.

# Preservation strategy: Pass the data to KM3NeT

*Decision by the ANTARES and KM3NeT collaboration:*

## **KM3NeT preserves and integrates ANTARES data**

→ Question of long-term preservation becomes task of the KM3NeT collaboration alongside their own data

→ need to manage transition

## **Transition made easy**

- Large overlap between ANTARES and KM3NeT researchers
- CCLyon as common HPC center
- Some overlapping infrastructure (e.g. Gitlab use)

# KM3NeT & ANTARES similarities

Using ANTARES data as KM3NeT “test” data

ANTARES data treated as example data in the KM3NeT Data Management Plan ([DMP](#))

## Where it helps

- *Open Science*: Converting to data format at Data Level 3 (science ready) and distribute
- *Data management*: Including large dataset to test with distributed access (Rucio)
- Providing & preserving *software*

Data volume to be transferred

- Raw data O(50 TB)
- Full standard processing (500+ TB)
- DSTs (100+ TB)
- Final legacy analysis data (1 TB)
- simulation: ~ x2
- Data base: ~ 2.5 TB

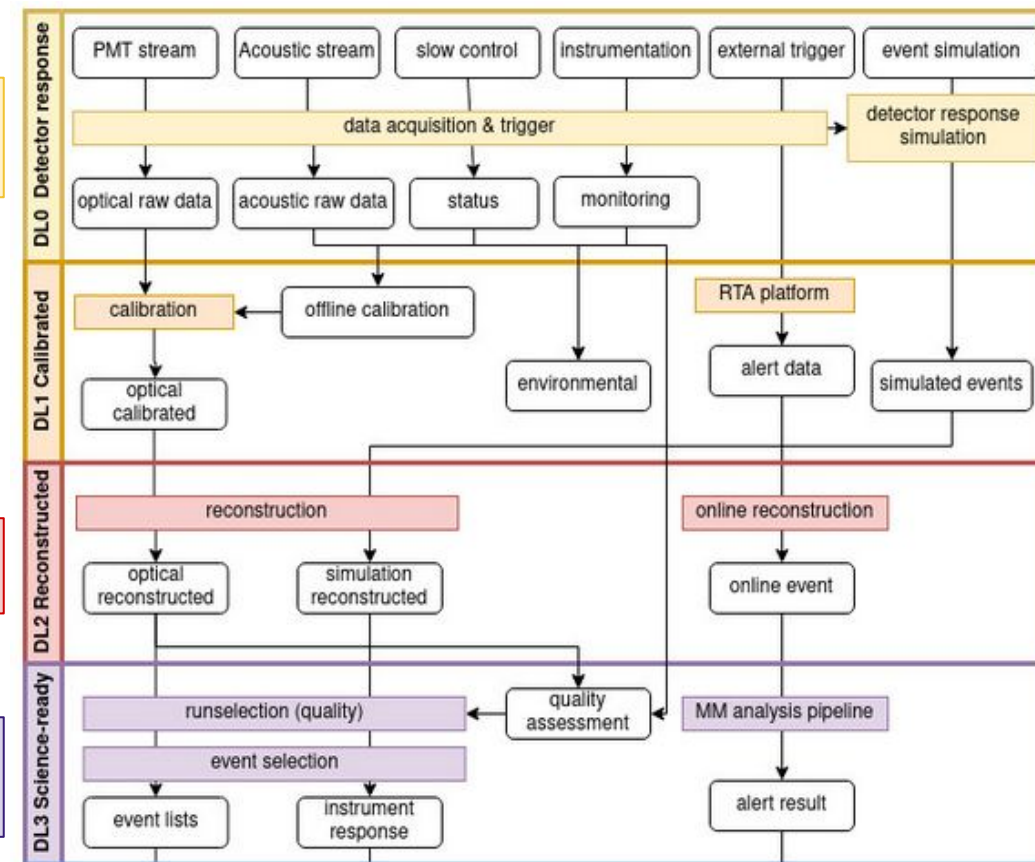
## ANTARES

Database & raw data

full events

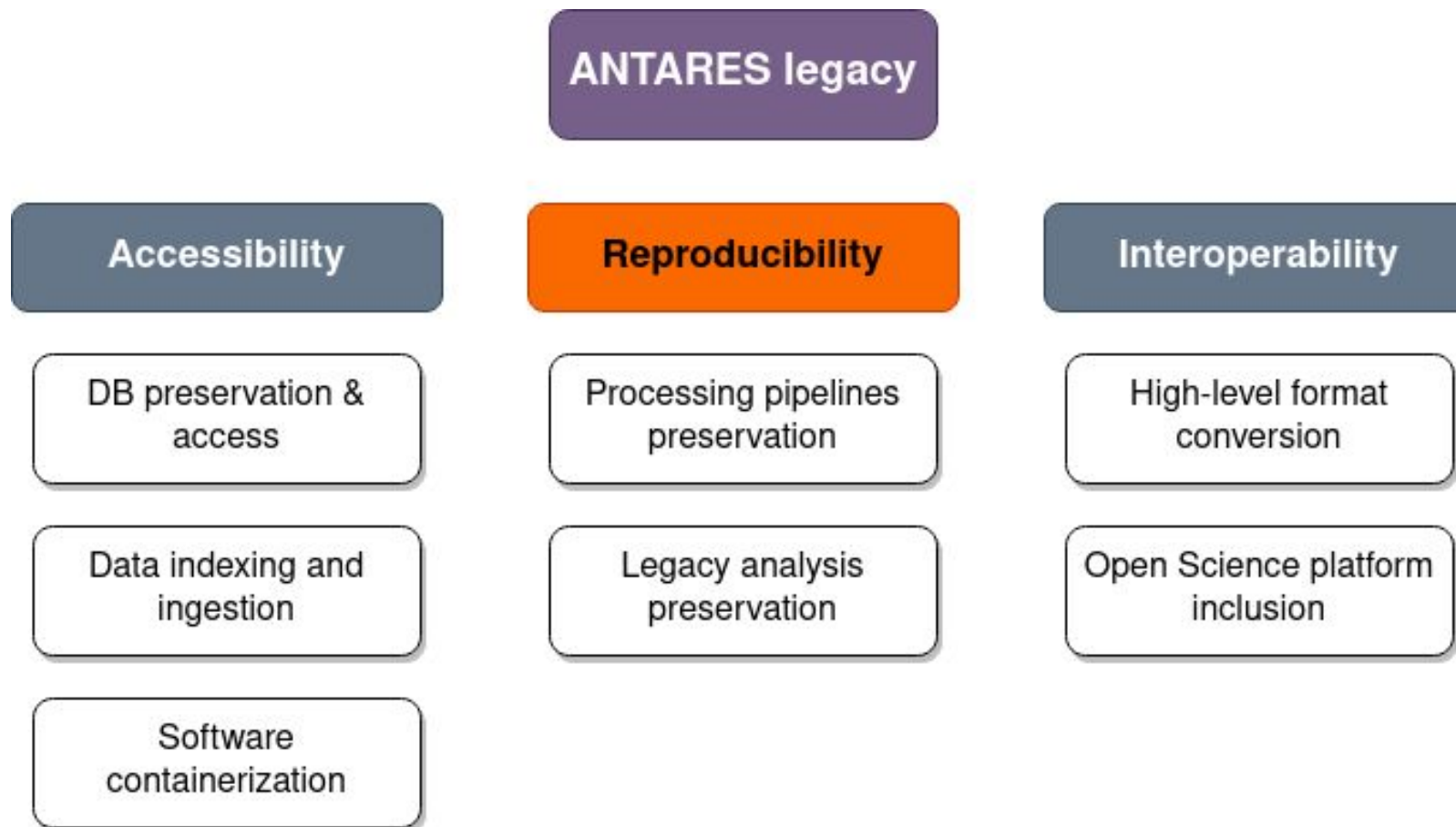
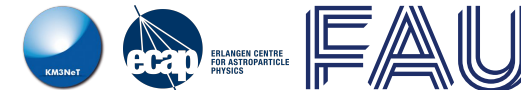
summary files

## KM3NeT lower data levels



# Goals of the preservation

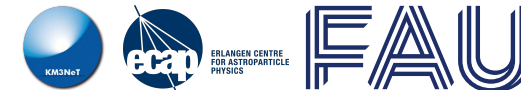
A rough overview for relevant tasks





# Detailing legacy tasks

How to make ANTARES data FAIR



## Database preservation & access

- Keeping DB accessible (Oracle update?)
- Considering containerization

## Data indexing and ingestion

- Ingest relevant data sets in KM3NeT DM system (Rucio)
- Make retrievable for use in relevant pipelines

## Software containerization

- Provide at minimum IO libraries for data reading/writing
- Preserve full processing pipeline as containers (singularity or similar)

## Processing pipelines preservation

Considering if we keep the *possibility to reprocess*

- containerize data processing pipeline
- containerize simulation chain

## Legacy analysis preservation

Providing analysis scripts: Requesting reproducible high-level analysis pipelines for currently ongoing “legacy” publications

## High-level format conversion

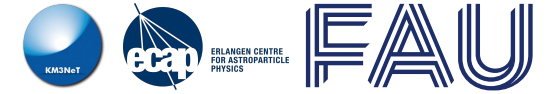
- Converter of ANTARES data format(s) to KM3NeT
- Starting from existing code

## Open Science platform inclusion

Provide usable science examples in open environment

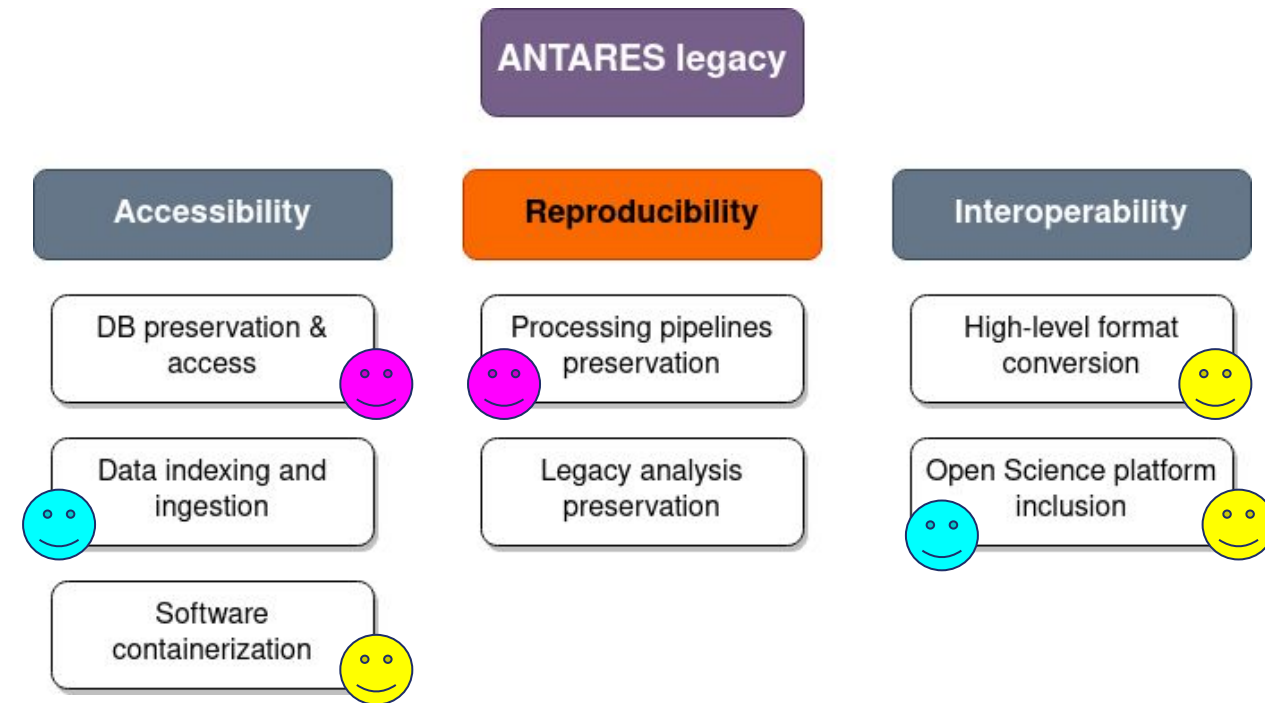
# Who is doing it

## Funding and opportunities - first considerations



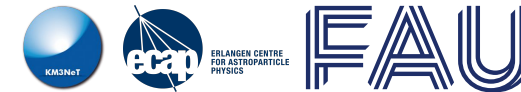
### Funding options

- Partially integratable in KM3NeT **infrastructure development (INFRADEV2)**
- Included in **ACME** call (HORIZON-INFRA-2023-SERV-01), currently starting
  - 4.2.10. *Access to neutrino data of ANTARES telescope.*
  - 4.3.2. *ANTARES and KM3NET neutrino telescope data analysis services*
- Still looking for **funding**



# Opportunities: Data management and metadata

Setting up distributed data access employing Rucio



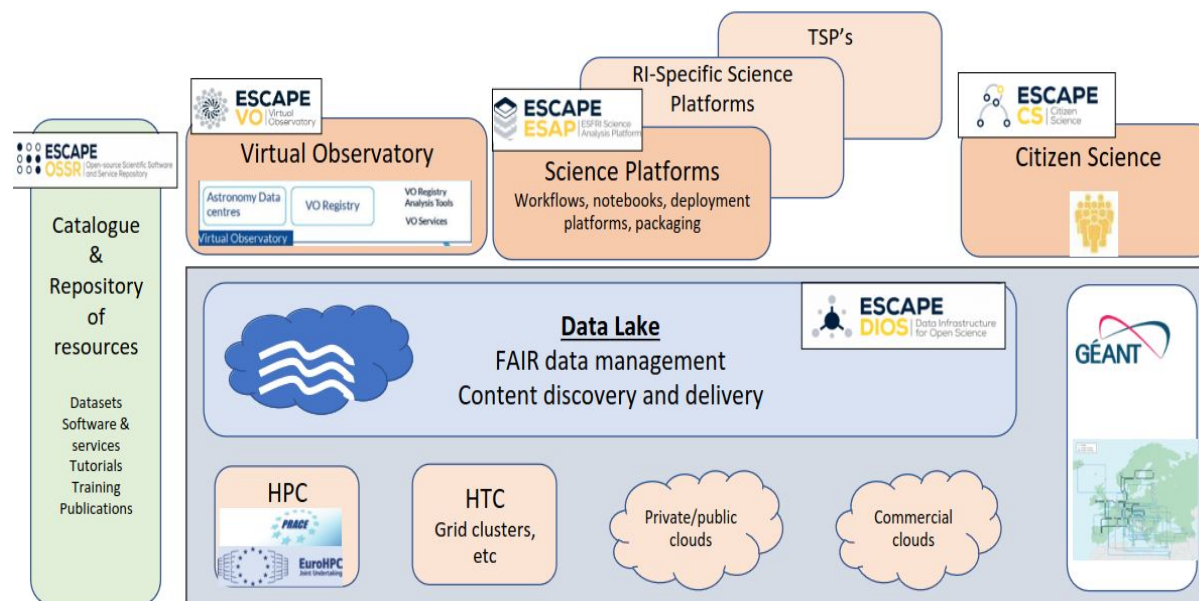
- Setting up KM3NeT data management
- Using Rucio for cross-cluster file management
- Still developing metadata schema etc.

→ Handling ANTARES legacy data as example case

→ further building on ESCAPE environment to then link to open infrastructure

## ESCAPE and European Open Science Cloud (EOSC)

- Developed common data lake, software repository & science platform
- Application used for further development of the Open Science System



© G. Lamanna, ESCAPE to the future

# Opportunities: Virtual Observatory

Integrating with the astronomy community

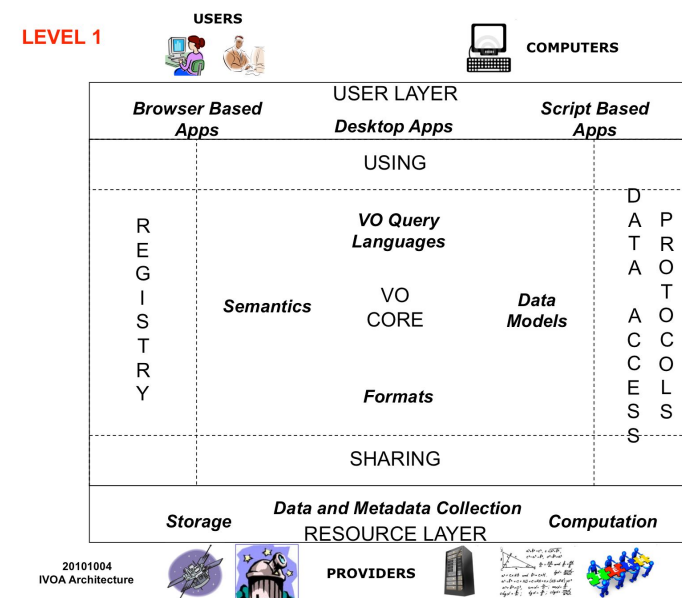
Decl	RA	Nhit	Beta	MJD
[deg]	[deg]	[deg]		[days]
19.5	68.2	21	1.0	54138.3105
-60.0	26.5	33	0.8	54138.5830
-29.8	82.1	34	0.3	54140.2299
-8.6	271.8	41	0.3	54140.6394
-32.3	261.4	45	0.5	54142.7042
-66.7	149.9	52	0.8	54159.4158



Name	Table Head	Description	Unit	UCD
ID	ID	Event ID	N/A	meta.id;meta.main
Decl	Dec (SI)	Declination	N/A	pos.eq.dec;meta.main
RA	RA (SI)	Right Ascension	N/A	pos.eq.ra;meta.main
Nhit	Nhits	Number of hits	N/A	meta.main
Beta	Beta	Angular resolution	N/A	meta.main
MJD	MJD	Epoch	N/A	meta.main

Example table data: ANTARES 2007-2017 event list @[vo.km3net.de](http://vo.km3net.de)

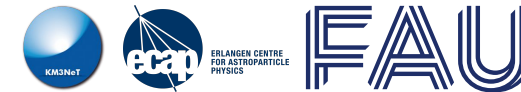
- VO highly standardized for astronomy analysis
- ANTARES event lists already provided
- Challenge: Protocols not made for low-count rate experiments → need to understand “extension”
- HEIG (High Energy Interest Group) forming @IVOA





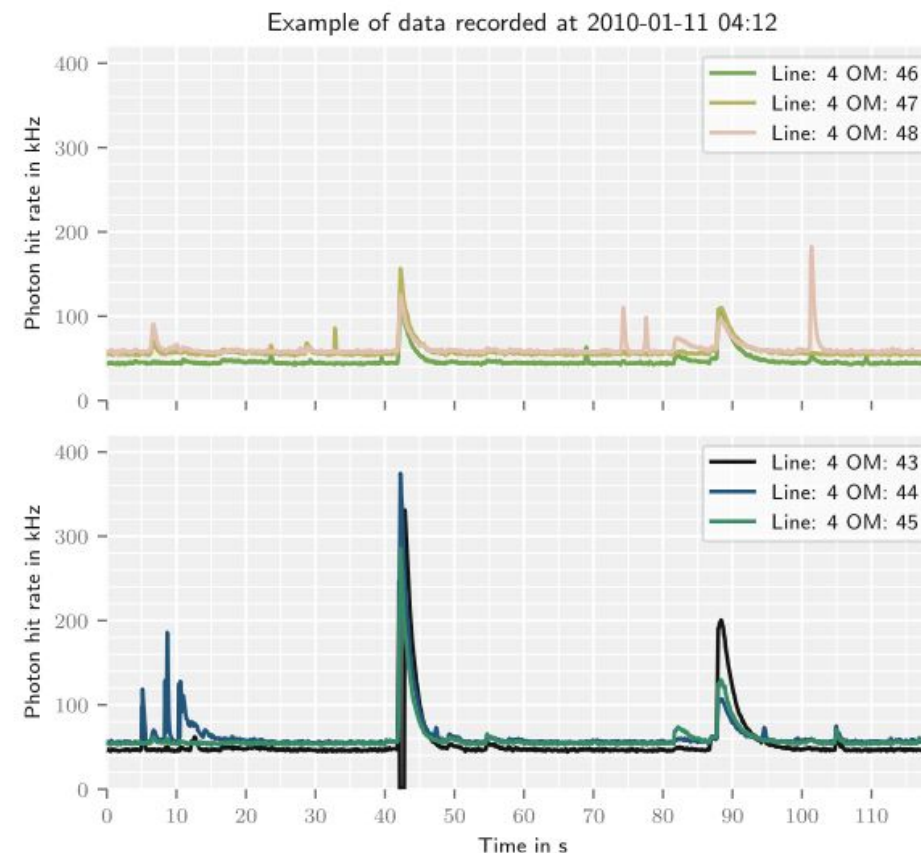
# Opportunities: Cross-domain research

## Deep-Sea data for Maritime research



### Database preservation & access

- Database contains valuable information on environmental conditions in the Deep Sea and bioluminescence rate
- ORACLE database hard to preserve (supported versions, licensing ...)
- Considering containerization
- Aiming to provide interface for Maritime Science: DEEPSEA project @OSCARS, not funded



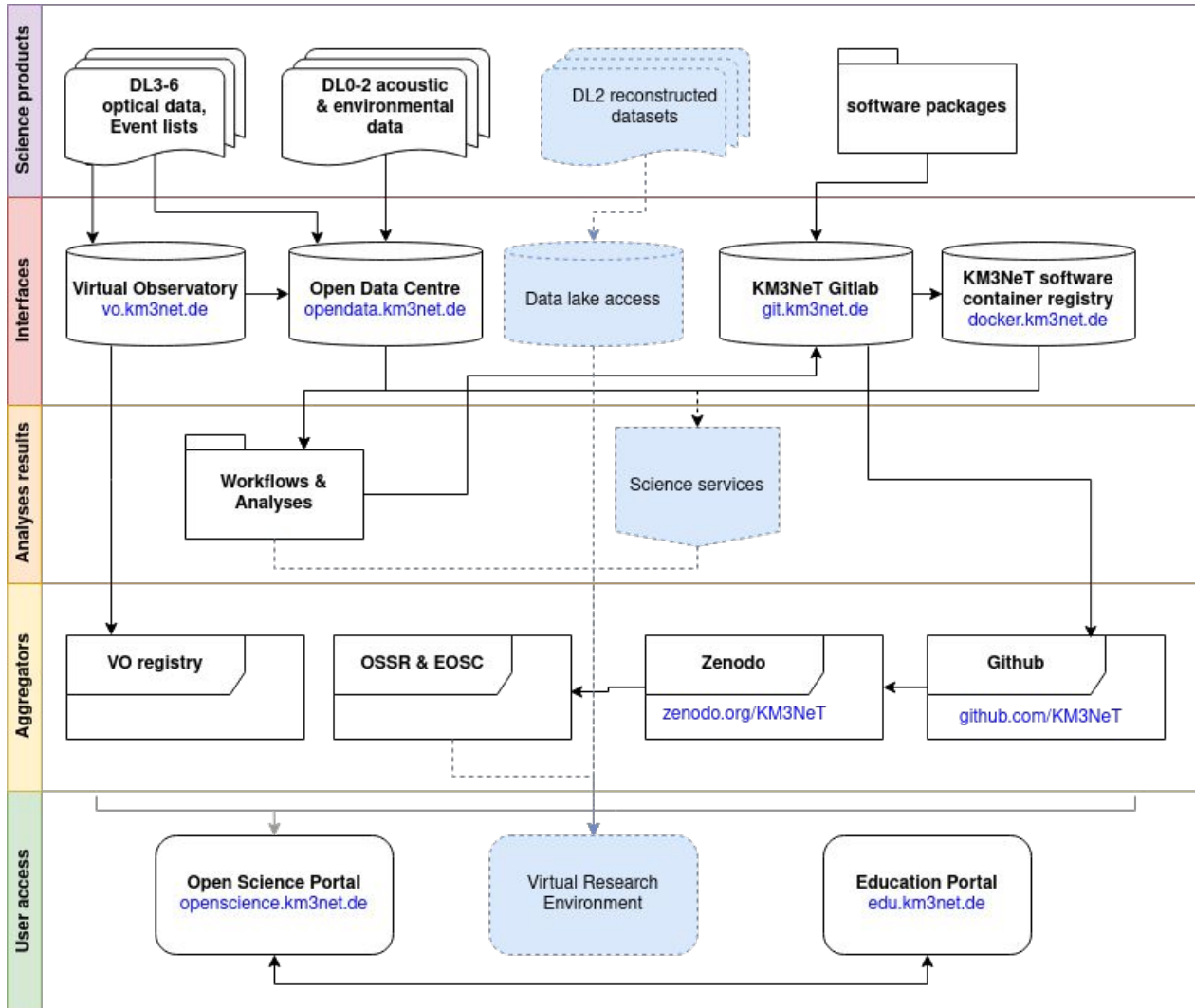
Studying bioluminescence flashes with the ANTARES deep-sea neutrino telescope. *Limnol Oceanogr Methods*, 21: 734-760. <https://doi.org/10.1002/lom3.10578>

- ANTARES data preservation will be facilitated by the KM3NeT collaboration
- Official commitments have been made by the collaborations, but implementation of the transitioning steps are still to be done
- ANTARES data will provide an excellent opportunity for KM3NeT to develop its interfaces for open science and data management
- Some funding still missing for essential parts of the transition



**Thank you for your attention  
&  
any questions?**

# Infrastructure: The KM3NeT Open Science System



- Defining data formats and standards for **science products**
  - Provide KM3NeT-side **interfaces**
    - For astrophysics: Virtual Observatory
    - For “everything”: Open Data Center
    - For software & Repositories: Gitlab and (docker) containers
  - Connection to **aggregators**: VO registry, Zenodo, Github, EOSC ...
  - Provide or integrate to **User platforms**
    - Open Science Portal, Education Portal
- Constant development & Improvement