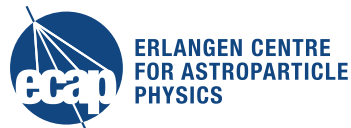


Computing in the KM3NeT Research Infrastructure

Jannik Hofestädt
VLVnT 2018 - Dubna
October 2018

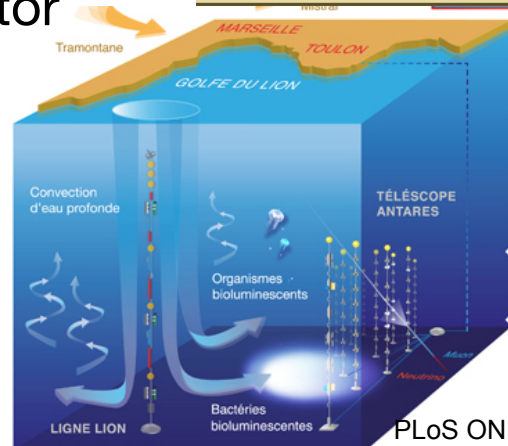
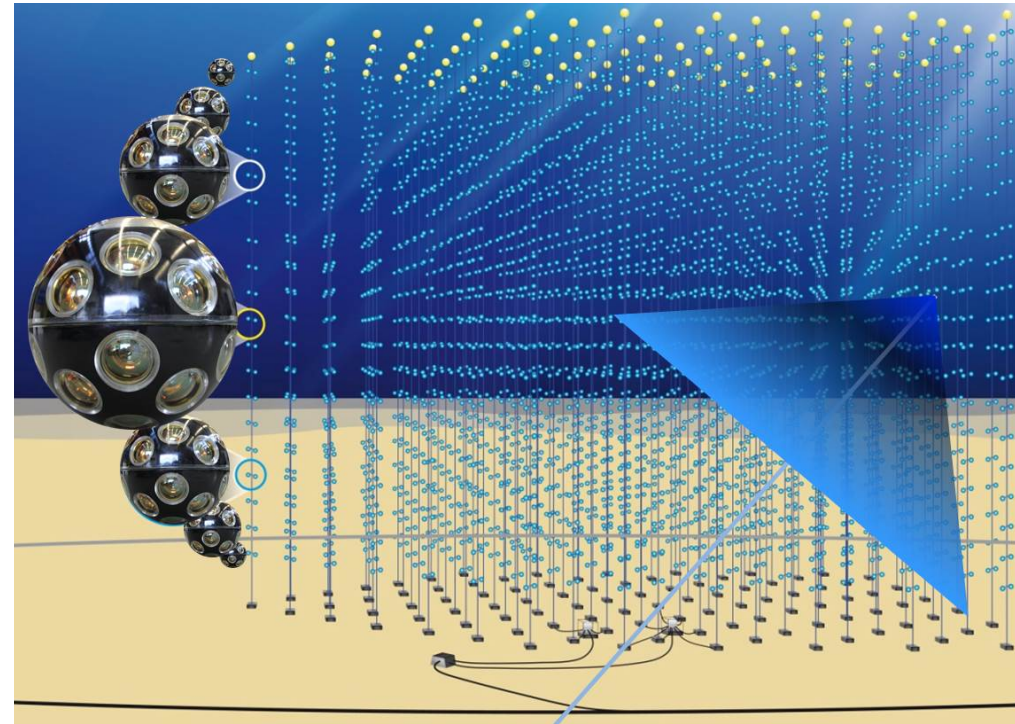
ecap



- Deep-sea research infrastructure in the Mediterranean Sea
- Includes:
 - ARCA: large telescope for neutrino astronomy/astrophysics
→ Italian site
 - ORCA: dense detector for neutrino physics
→ French site
 - Nodes for Earth and Marine sciences

↑ Talk by P. Coyle

→ Talk by D. Lefevre



PLoS ONE 8 (7) 2013

Detector

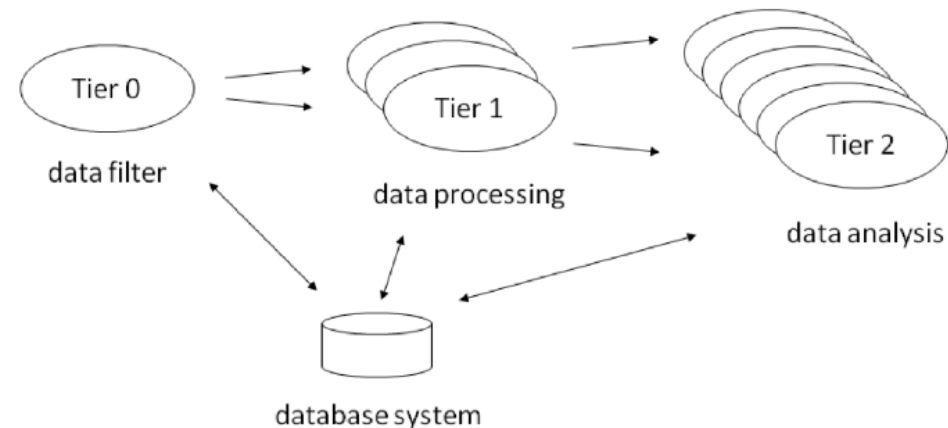
- 31 3-inch PMTs = 1 digital optical module (DOM)
- 18 DOMs = 1 detection unit (DU)
- 115 DUs = 1 building block (BB)
- 3 BBs = KM3NeT Phase 2

Altogether: 192500 PMTs, 5-10 kHz single-photon rate each,
all data to shore → some GB/second

Data processing

- Online data filter (on shore)
- Calibration and reconstruction
- Data analysis

- + Monte Carlo simulations
- + Online analysis for alerts etc.
- + Earth and Sea sciences



Requires advanced
data management



- Computing model
- FAIR data:
 - findable
 - accessible
 - interoperable
 - re-usable



KM3NeT INFRADEV – H2020 – 739560

KM3NeT Data Management Plan

KM3NeT-INFRADEV GA DELIVERABLE: D4.1

Document identifier:	KM3NeT-InfraDev-WP4-D4.1_v1.4
Date:	22 June 2017
Work package:	WP4
Lead partner:	FAU
Document status:	Endorsed by PMB and KM3NeT IB
Dissemination level:	Public



- Tier-like structure, mixed access: GRID + direct (batch)

Tier	Computing Facility	Processing steps	Access
Tier-0	at detector sites	triggering, online-calibration, quasi-online reconstruction	direct access, direct processing
Tier-1	computing centres	calibration, reconstruction, simulation	direct access, batch processing and/or grid access
Tier-2	local computing clusters	development, simulation, analysis	varying

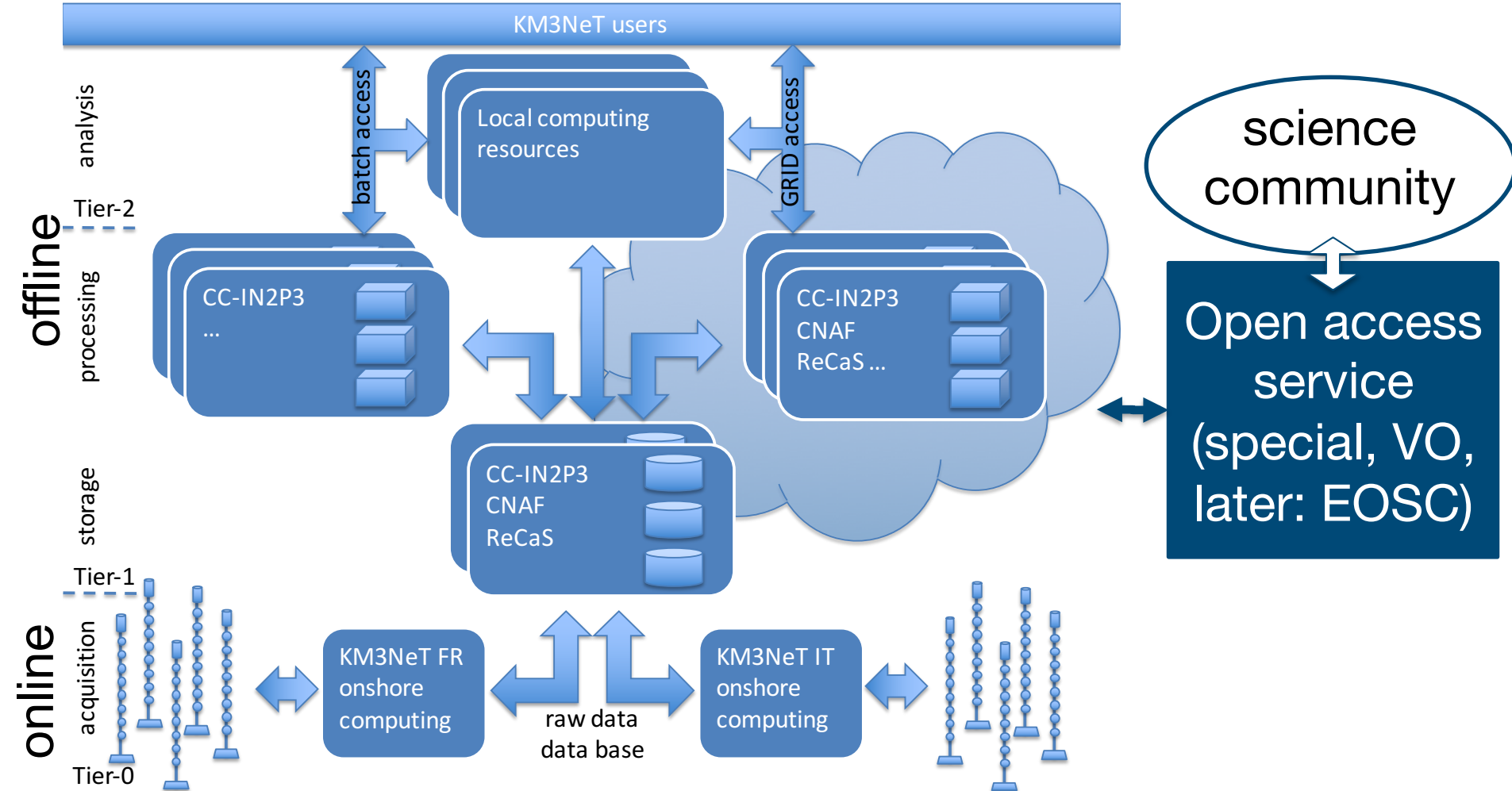
- Data transfer between the computing centers based on GRID access tools (where applicable)
- Central services funded through CNRS and INFN, additional services by the collaboration institutes

KM3NeT Data Management



Internal:

External:



1 building block, 1 year



processing stage	size per proc. (TB)	time per proc. (HS06.h)	size per year (TB)	time per year (HS06.h)	periodicity (per year)
Raw Data					
Raw Filtered Data	300	-	300	-	1
Monitoring and Minimum Bias Data	150	-	150	-	1
Experimental Data Processing					
Calibration + intermediate steps (incl. Raw Data)	750	24 M	1500	48 M	2
Reconstructed Data	150	119 M	300	238 M	2
DST	75	30 M	150	60 M	2
Simulation Data Processing					
Air showers	100	14 M	50	7 M	0.5
Atm. Muons	50	1 M	25	638 k	0.5
Neutrinos	2	22 k	20	220 k	10
Total:	827	188 M	995	353 M	

KM3NeT Preliminary



- Central goal: prompt dissemination of scientific results, new methods and implementations
- Data formats and meta-data following common practice (root, hdf5, fits, xml, ascii) → allow for integration in eCommons
- Use of existing eCommons (e.g. from ANTARES) and specific development via ASTERICS and KM3NeT-INFRADEV
- Data storage:
 - Essential: reproducibility of all scientific results and data usability over full lifetime of experiment (>10 years after end of operation)
 - Parallel storage of low- and high-level data at CC-Lyon & CNAF (long-term commitments, pledged resources)
 - Central services like software repository, central software builds
- Data processing by specialized service group
- Data access: via WAN/GRID access tools (xrootd, iRODS & gridFTP)



- Public access to data
 - Summary data (event information plus quality information) after fixed latency (typically 2 years)
 - Web-based downloads of data and softwares
 - Also: Simulation data
- On request: More (detailed) data, earlier releases, etc.
- Harmonise with wider community (→ GNN and ASTERICS), use common tools/platforms (→ Virtual Observatory), link to other eCommons (e.g. EOSC)

ASTERICS = Astronomy ESFRI and Research Infrastructure Cluster;
EOSC = European Open Science Cloud;
GNN = Global Neutrino Network.



- KM3NeT construction started, detector with 3 blocks by ~2023
- Data management plan and computing model established
- KM3NeT adopts tier-model for data management (large data volumes, significant CPU requirements)
- Open data access in preparation (includes data archiving)
- KM3NeT profit strongly from GNN harmonisation