Friedrich-Alexander-Universität
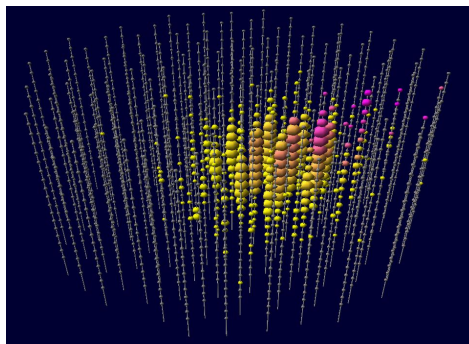Erlangen-Nürnberg

Open Science in KM3NeT

Jutta Schnabel

KM3NeT Town Hall Meeting 2022

Catania, 20th September 2022

Open Science as Collaboration goal
## The KM3NeT collaboration

KM3NeT is

- currently under construction - we are starting taking data

- at the same time building the computing infrastructure - to **use and share** data

Open Science Policy

KM3NeT supports the aims of open data and open science and commits to implement the necessary steps wherever possible. This includes open access data supporting publications, open source software and open data in general including the information needed to appropriately use the data.

The Open Science Paradigm
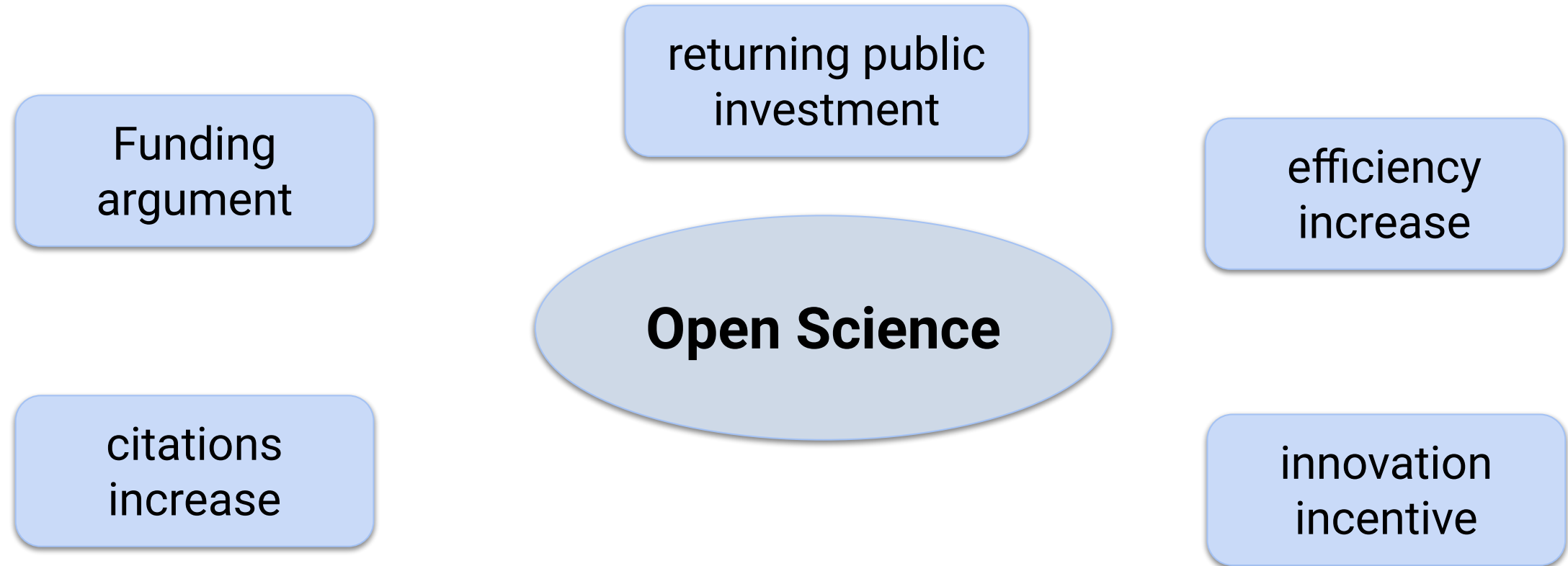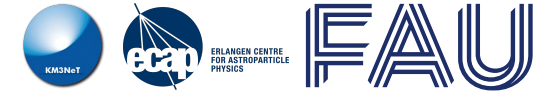
What does it actually mean?
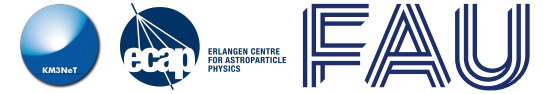
Open science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, amateur or professional.

Open science is transparent and accessible knowledge that is shared and developed through collaborative networks.

Wikipedia (Open Science)
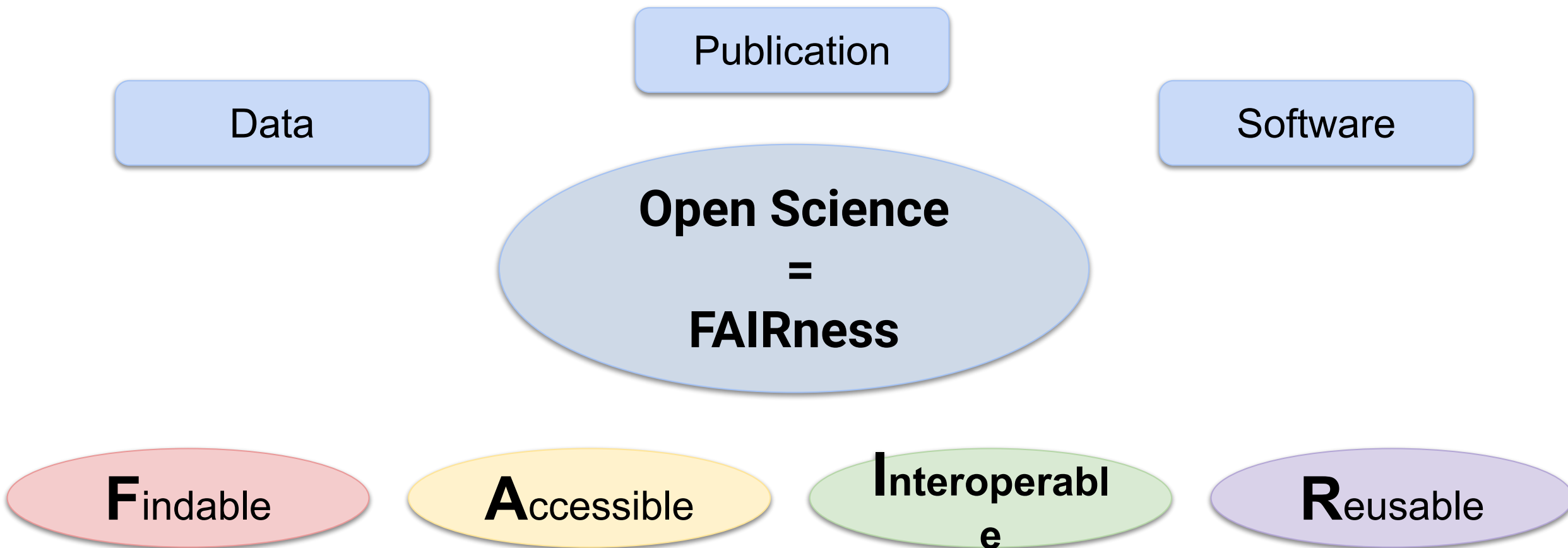
Why care about Open Science?

Not only out of good will

returning public investment

Funding argument

efficiency increase

Open Science

citations increase

innovation incentive

OECD (2015-10-15), "Making Open Science a Reality", OECD Science, Technology and Industry Policy Papers, No. 25, OECD Publishing, Paris. http://dx.doi.org/10.1787/5jrs2f963zs1-en

Publication

Data

Software

Open Science
=
FAIRness

Findable

Accessible

Interoperable

FReusable

Using large open platforms

Findable

Metadata, Identifiers, Archiving

Data

Publication

Software

arXiv.org

Integrating in different environments

Data always goes along with software
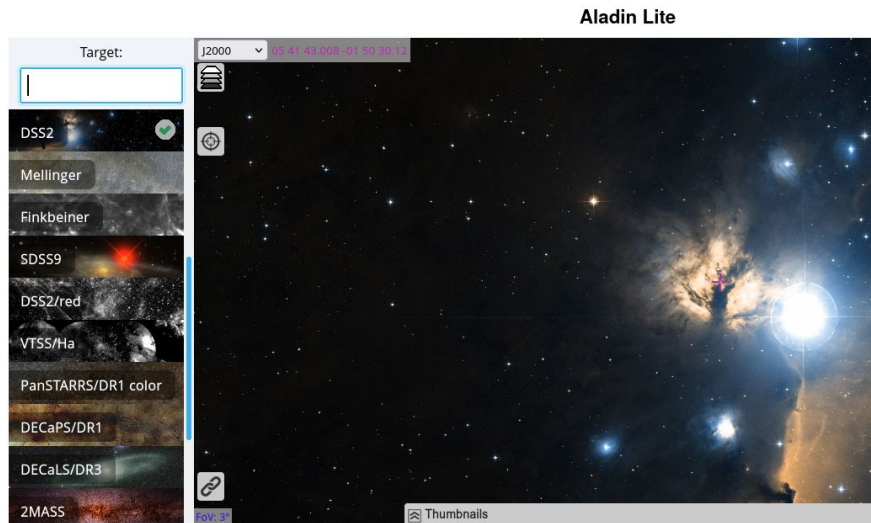
Data

Software

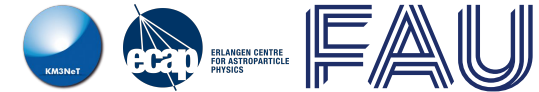Aladin Lite

jupyter   docker

analysis environments

containerization

notebooks

Standardizing data formats and use
Either use what is there or commonly produce your own

Interoperable

Data

Software

Documents & Standards

DOCUMENTS   XML SCHEMA   VOCABULARIES   DOC SUBMISSION

- *Technical Specifications*
- *Notes*
- *Promotion process*
- *IVOA Technical Assessment and Roadmap Documents*
- *Submission Log*

Technical Specifications
> >

Group	Title	Most stable	In progress	Version history
App	SAMP - Simple Application Messaging Protocol	1.3		1.3 1.3 1.3 1.3 1.3 1.2 1.2 1.2 1.11 1.11 1.10 1.00
	VOTable - VOTable Format Definition	1.4		1.4 1.4 1.4 1.4 1.4 1.4 1.3 1.3 1.3 1.2 1.2 1.2 1.20 1.20 1.10 1.00
	MOC - HEALPix Multi-Order Coverage Map	1.1	RFC	2.0 2.0 2.0 2.0 2.0 2.0 1.1 1.1 1.1 1.1 1.1 1.0 1.0 1.0 1.0 1.0

γπ A **Python** package for **gamma-ray** astronomy

Defining schemas

providing converters

Standards to publications

A publication is more than the paper

Publication

Data

Software
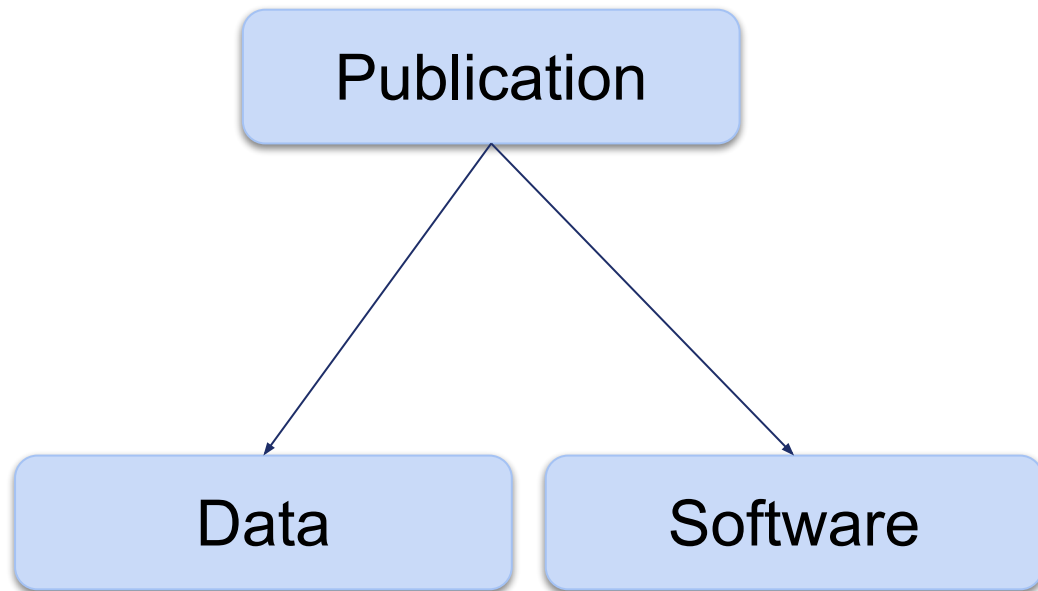
Analysis repositories

+   documentation!

Welcome to KM3NeT Science!
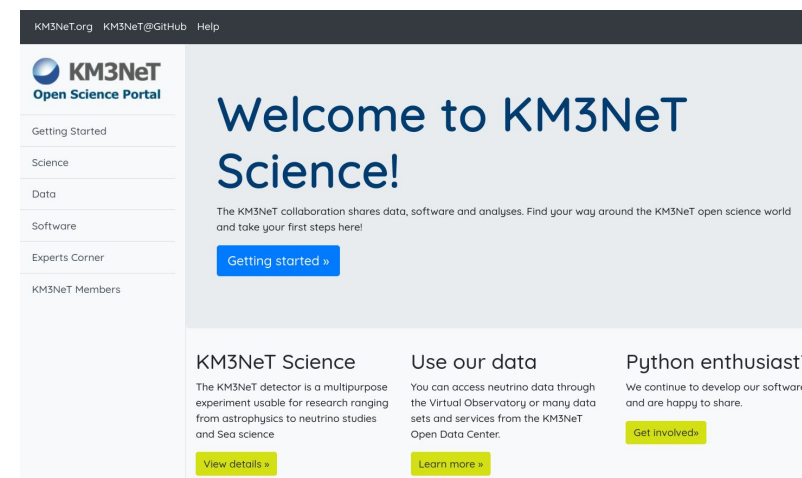
Open Science needs change
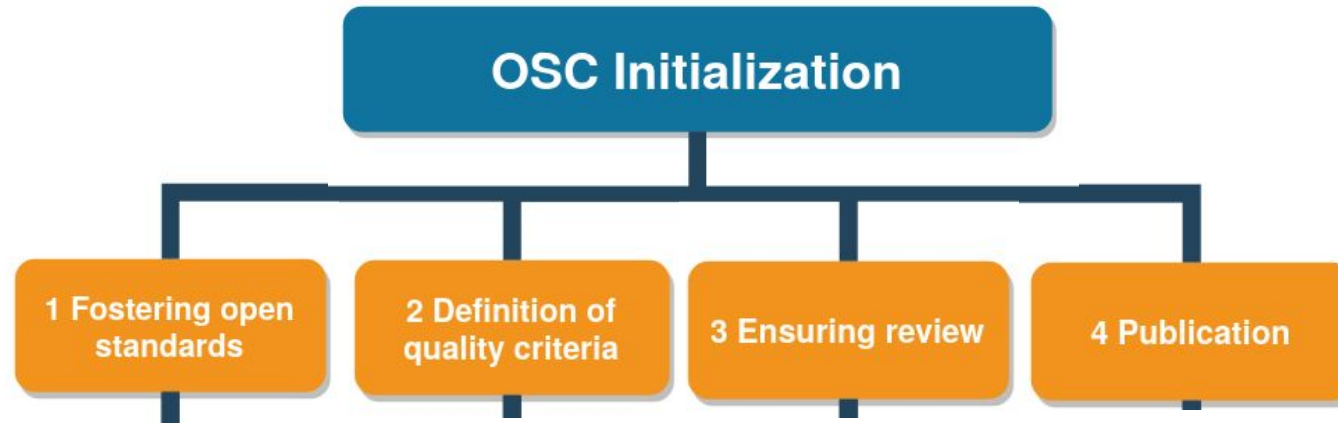
Internal implementation and lobbying

The Open Science Committee

Implementation in the collaboration needs

- Discussions
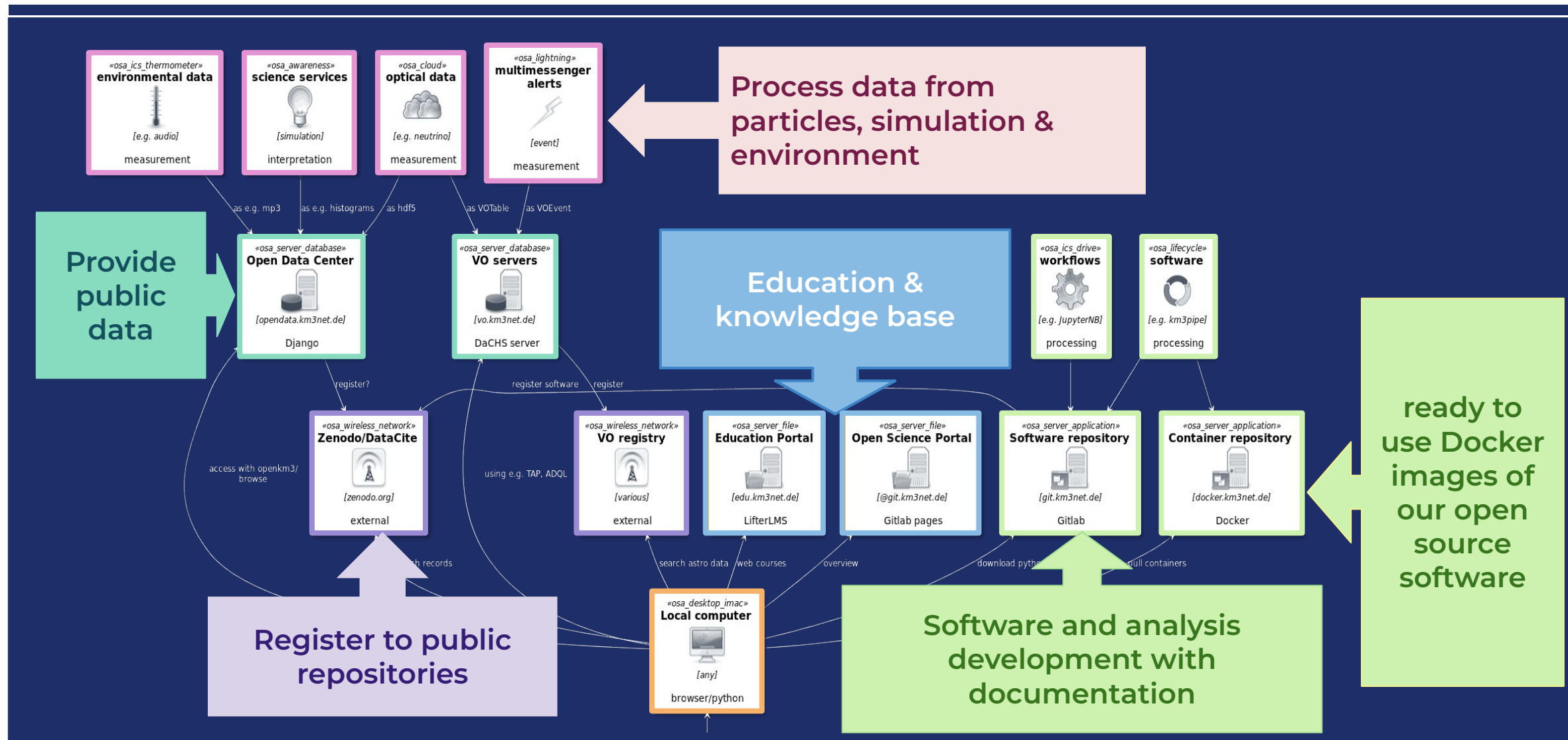
- Development

- Standardization & Guidelines

- Training

Starting with prototyping

In data, software, services with the KM3NeT Open Science environment
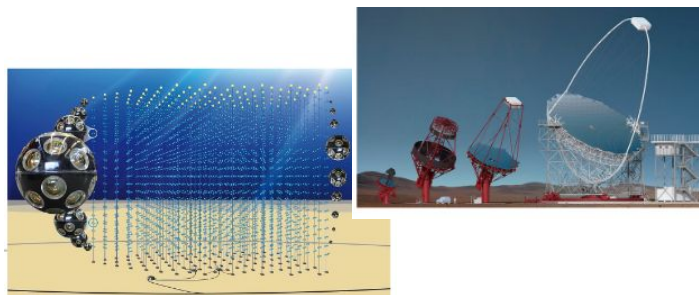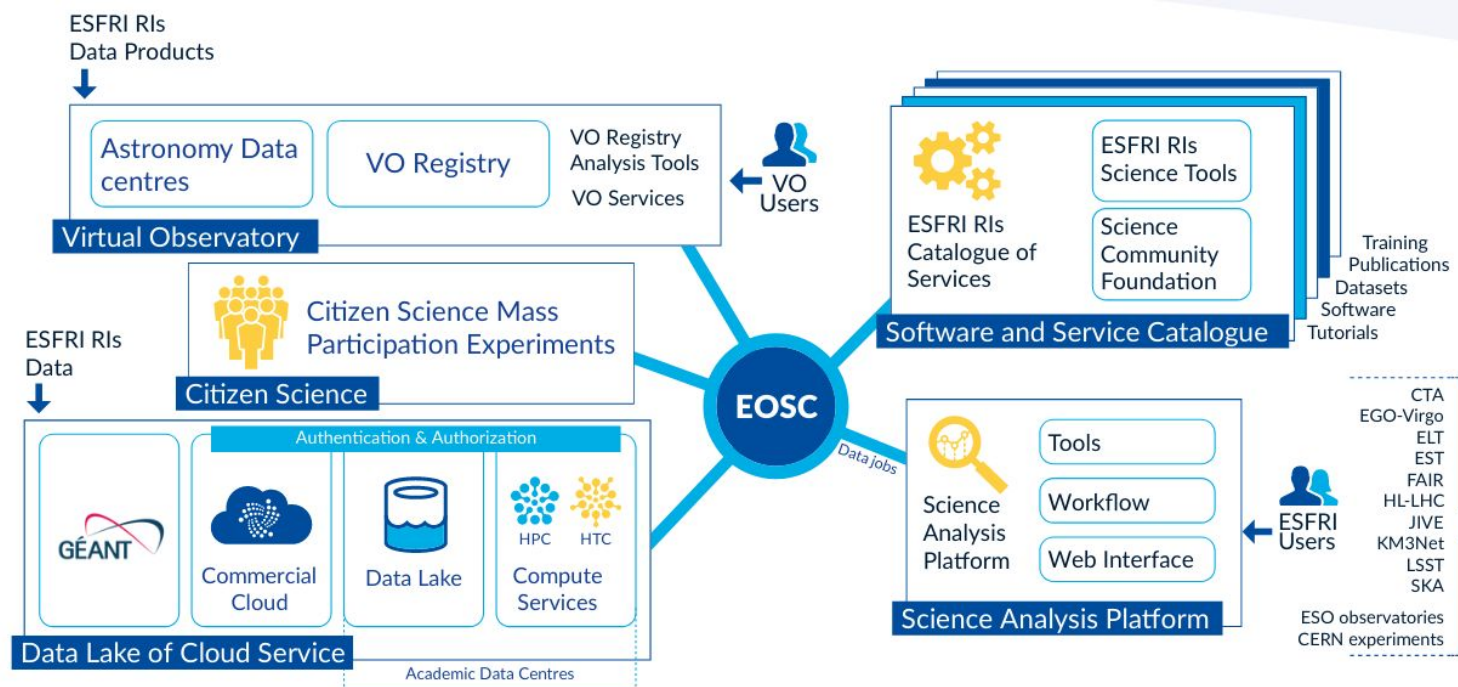
Developing with common initiatives

From institute initiatives to large-scale cooperation

Development happens in the community

- in ESCAPE/EOSC future
  - data lake infrastructure
  - common analysis platform
  - software sharing
  - common formats
- in various multi-messenger initiatives

Current tools & developments

Starting points for open data

- Virtual observatory
    - hoping to host ANTARES data
    - using VO format for alerts
    - initiative to develop further metadata with IVOA (common with CTA)
- Common VHE format
    - IRFs
    - integrating with gammapy
    - common GW initiative
- in particle physics: cooperation on computing level
    - HPC computing tools: DIRAC and RUCIO for data processing
    - scientific workflows (containerization of software, jupyter environments)
    - further interaction in machine learning, oscillation analysis etc. yet terra incognita

What is KM3NeT data?

And what is relevant for physics?

Particle event data

- Level 2 data: fully calibrated events, including hit information (ROOT files)

- Level 3 data: fully reconstructed event without hit information (ROOT, hdf5)

- Level 4: Selected data sets, single events (VOTable, VOEvent)

Event simulation

- Signal (cosmic neutrino) and background (atmospheric) events
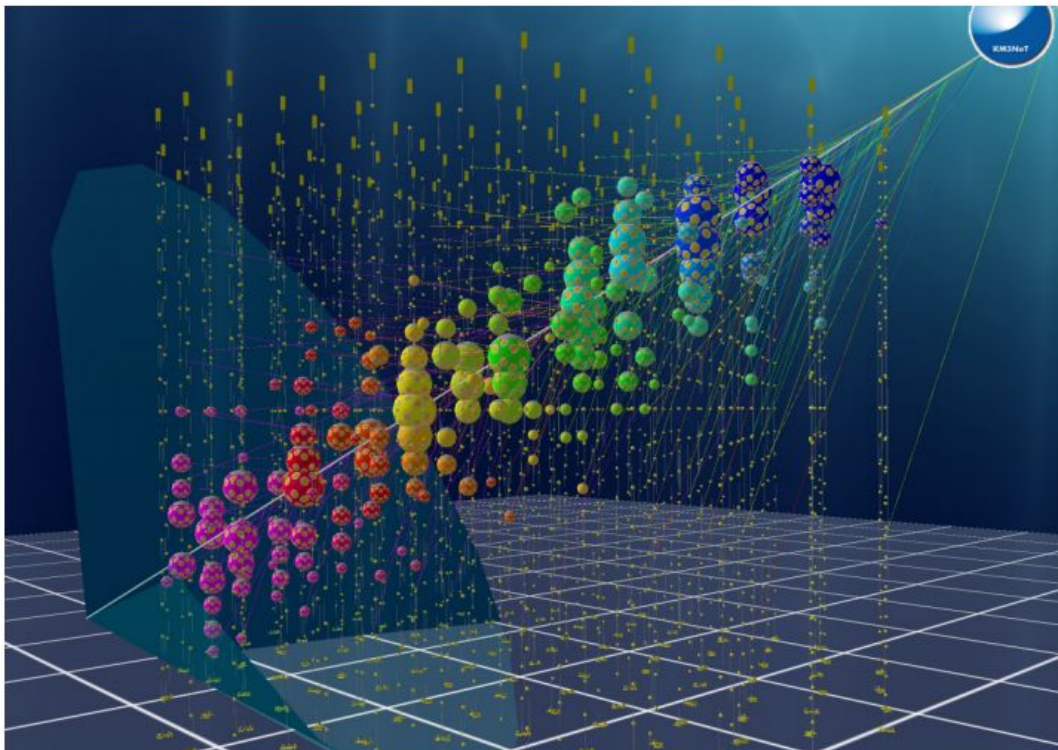- **Analogous processing** to measurements
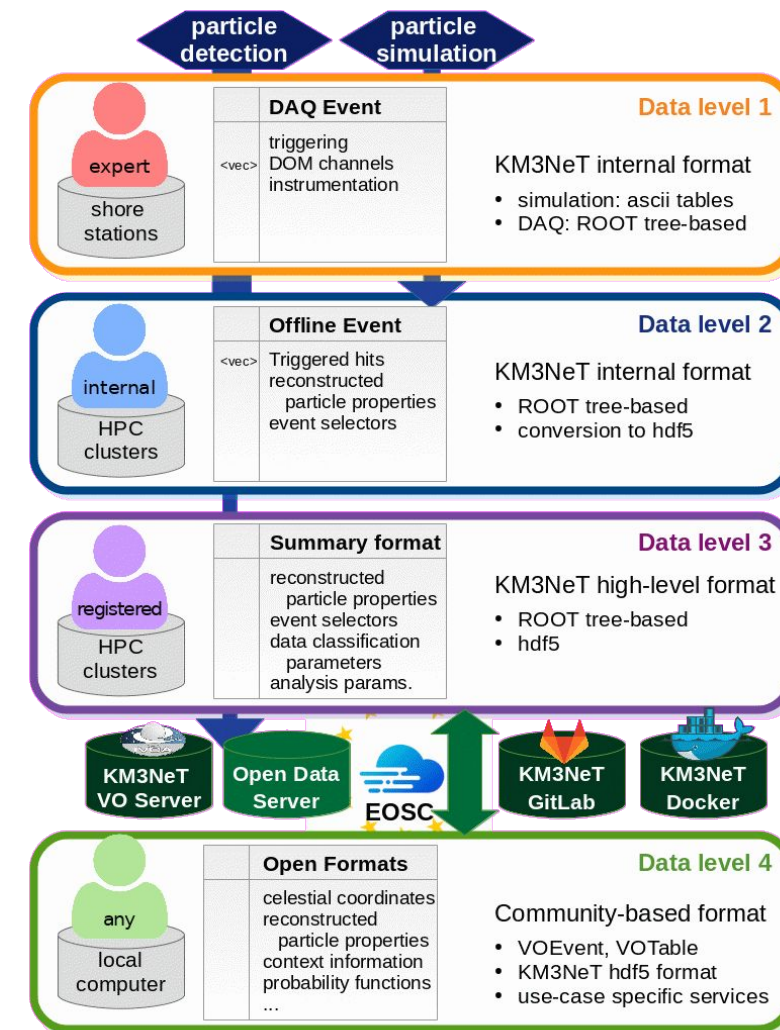
"Easy" data: neutrino events

The main output of KM3NeT

„Full" event (i.e. particle detection!)

event identification	detector status	<photon detections x̄, t, A>

„Reduced" event

reconstructed particle properties			direction time energy, resolution ...	

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
-13.0	93.6	25	0.7	54160.4830
-26.2	266.7	28	0.8	54160.6180
23.5	121.7	41	0.5	54161.4361
-70.7	47.1	30	0.9	54165.5838
-55.0	284.4	36	0.5	54169.0685

Example files of KM3NeT ROOT files:
https://github.com/KM3NeT/km3net-testdata

Neutrino events in astrophysics

Sharing in the Virtual Observatory
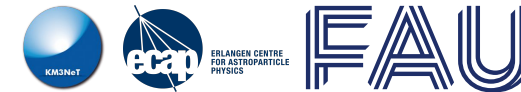
- Running server with DaCHS software

- Registered as data provider to the VO
 (ivo://km3net.org)

- Can publish data sets to the VO registry - done
  with Antares 2007-2017 data

- Data accessible through widely used tools in
  astrophysics (Aladin, Topcat …)

Virtual observatory server
http://vo.km3net.de/

Information on resource 'ANTARES 2007-2017'

Neutrino candidates from full-sky search from 2007-2017

Services defined within this resource descriptor

- ANTARES 2007-2017

Tables defined within this resource descriptor

- ant20_01.main – queriable through TAP and ADQL

[Manage RD]

Please report errors and problems to the site operators. Thanks.

KM3NeT

Help

Service info

Metadata

Description

Neutrino candidates from full-sky 2007-2017
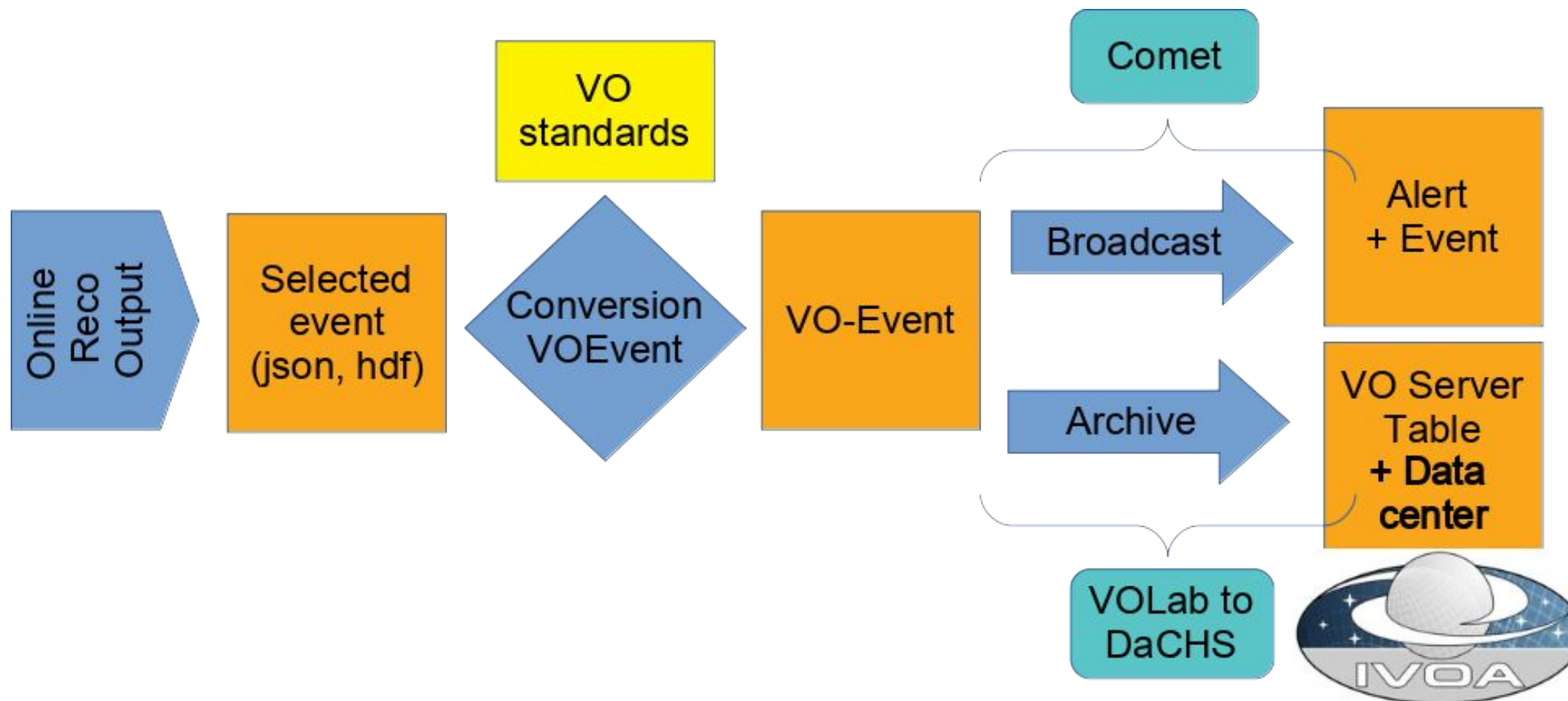
Keywords

neutrino

Creator
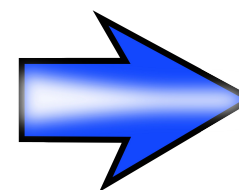
Integration of further (and legacy) ANTARES data in KM3NeT environment planned

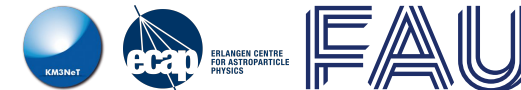# Single alert events

Different form and pipeline for VOEvents

- Converting alert info (json) to VOEvent
- Access additionally provided through **data center**

MM VO Event
https://git.km3net.de/j
schnabel/kmeta/-/tree/
master/examples/appl
ications/VOEvent

Events outside of astrophysics

Dealing with less standardization in particle physics

Introducing: The KM3NeT Open Data Center

- For all data not publishable through the IVOA, serving as interface and/or server to the data

- Including also link to data sets on VO server

- Based on Django REST API

- Usable for event data sets (hdf5-files with standardized metadata), plots or services, environmental data ...

- Data accessible through webpage, through REST-API or python based package (openkm3)

Open Data Center
http://opendata.km3net.de

KM3NeT

Open Data Center

We make our data available for you!

You can find a description of how to use our data at the

Current uploads

KM3NeT test data

ORCA 4-line events

More Info

How to add relevant probability functions?

Adding instrument response, background and other auxiliary simulated information

No VO standard for neutrino regime - starting our own

- extended header with content identifier
- tabulated data (e.g. bin content, function values …)
- ready for use in python environment

Example: Effective area for ANTARES 2007-17 Point Source search

```
ks.print_index()

annotated_aeff0
===============
header:
    name:            ANTARES 2007-2017 effective area
    description:     effective area for E-2 source sp
    contact:         antares.spokesperson@in2p3.fr
    instrument:      ANTARES
    license:         Creative Commons 4.0 Internation
    reference:       https://antares.in2p3.fr/public
```

```
plain = ks.get("annotated_aeff0") # not so nice, make i
plain.data
```

```
{'Log(E1/GeV)': [1.0,
  1.2,
  1.4,
  1.6,
  1.8,
  2.0,
  2.2,
  2.4,
  2.6,
  2.8,
  3.0,
  3.2,
  3.4,
  3.6,
```

get as table

```
table = ks.get("annotated_aeff0", loadoption = "pandas")
table.data
```

	Log(E1/GeV)	Log(E2/GeV)	AEFF/m2
0	1.0	1.2	5.153940e-12
1	1.2	1.4	1.614220e-10
2	1.4	1.6	4.891220e-09
3	1.6	1.8	1.245090e-07
4	1.8	2.0	9.157530e-07
5	2.0	2.2	4.372520e-06

get as dataframe

```
nice = ks.get("annotated_aeff0", loadoption= "plot")
```


ANTARES 2007-2017 effective area

get as plot
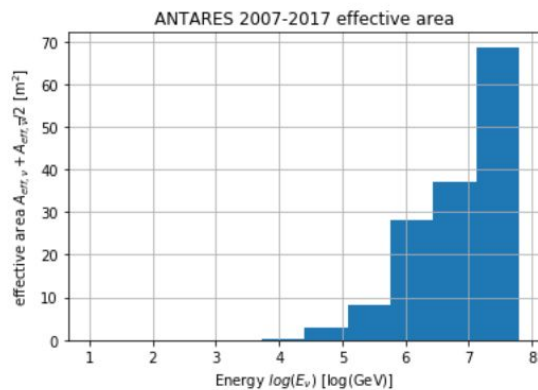
```
table.get_origin()
```

```
{'range': {'time': {'measurement_start': datetime.datetime(2007, 1,
  'measurement_stop': datetime.datetime(2018, 1, 1, 0, 0),
  'ktype': 'tbd'},
  'zenith': {'min': -45, 'max': 0, 'ktype': 'tbd'}}}
```

```
table.get_paraminfo()
```

```
{'lower_edge': {'columnname': 'Log(E1/GeV)',
  'name': 'Energy',
  'description': 'logarithmic reconstructed energy of the neutrino',
  'unit': 'log(GeV)',
  'symbol': 'log(E_{\\nu})'},
  'upper_edge': {'columnname': 'Log(E2/GeV)',
  'name': 'Energy',
  'description': 'logarithmic reconstructed energy of the neutrino',
  'unit': 'log(GeV)',
  'symbol': 'log(E_{\\nu})'},
```
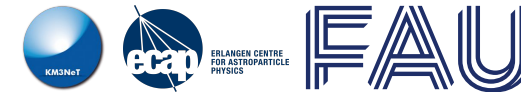
get metadata

Environment integration
Making it all python accessible

pip install git+https://git.km3net.de/open-data/openkm3

- ● dependency: pyvo for VO interface
- ● "KM3Store" as access point for all data + services

```python
from openkm3.openio import KM3Store
ks = KM3Store()
```

```python
table = ks.get("one_week_orca", ["events"], "pandas")
table.data
```

	angular_error	azimuth	dirz	energy	internalID
0	0.004341	5.108108	-0.968124	22.017775	km3net.44.61!
1	0.001181	3.358250	-0.990520	128.639694	km3net.44.61!
2	0.003534	6.146142	-0.744035	24.362326	km3net.44.61!
3	0.008247	2.954967	-0.784588	15.714084	km3net.44.61!
4	0.006967	5.618736	-0.529338	137.616933	km3net.44.61!

- ● entries are data sets, single files or services
- ● access logic coded in package

```python
ks.print_index(include_technical=True)
```

```
one_week_orca
=============
tables: ['events', 'group_info', 'header']
header:
        author:                 b'The KM3NeT collaboration'
        contact:                b'opendata@km3net.de'
        instrument:             b'ORCA'
        license:                b'Creative Commons 4.0 International'
        measurement_start:      b'2019-10-03T06:00:00.544000000'
        measurement_stop:       b'2019-09-17T06:00:00.202000000'
        reference:              b'http://www.km3net.org/'
url:    http://vo.km3net.de:82/storage/one_week_orca.h5
type:   application/x-hdf5
local:  /home/jutta/Desktop/openkm3/examples/orca_data/.openkm3/one_week
```

Beyond the self-made approach
Initiative for a Very-high-energy Open Data Format

The Gamma-ray Astronomy Data Formats Initiative (GADF)

technische universität dortmund

- Ground-based gamma-ray observatories are mostly run by particle physicists as closed collaborations
  - No or very limited public data access
  - Mostly proprietary analysis software
  - Custom data formats, usually strongly coupled with the proprietary software
  - Mostly based on CERN's ROOT
- Satellites and the Cherenkov Telescope Array (will) operate as open observatories
- The current generation of telescopes want a better archival format
- Strong scientific use-cases for multi-instrument analysis
- ⇒ Effort for a common, software-independent data format
- ⇒ Development of Open Source analysis software ("science tools")

Community Initiative for a VHE Open Data Format
Maximilian Nöthe and Lars Mohrmann for the GADF Initiative
DPG Spring Meeting 2022 – Heidelberg

2 / 12

Members of the Initiative

- Follow-up initiative of the GADF since this year: VODF

- Aiming to provide IRFs and Event lists in multiple contexts (including VO)

- aiming for further use of gammapy and common developments
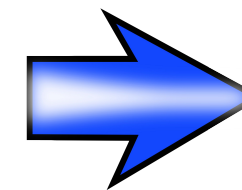
converters to VODF
easy use of gammapy
integration into VO

Data processing and access

Using the strength of particle physics community - sharing technology

- CERN-developed software dedicated to high-volume data processing
  - [Rucio](#) for scientific data management
  - exploring the [Dirac interware](#) for job management
  - exploring workflow management software ([snakemake,](#) nextflow)

- projects for development & integration: ESCAPE

ESCAPE DIOS
X. Espinal, ESCAPE
2nd Review

ESCAPE ESAP
J. Swinbank, ESCAPE
2nd Review
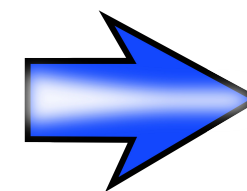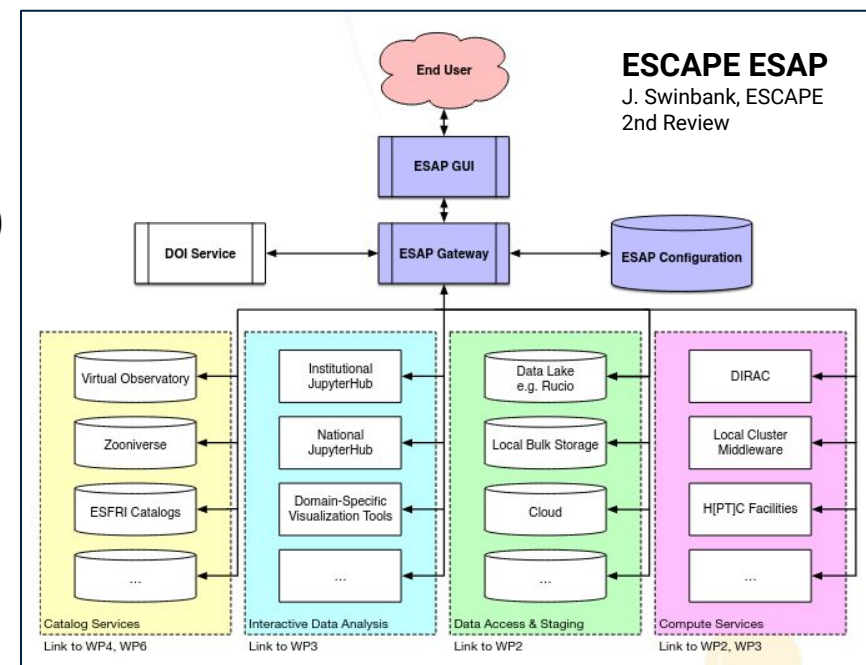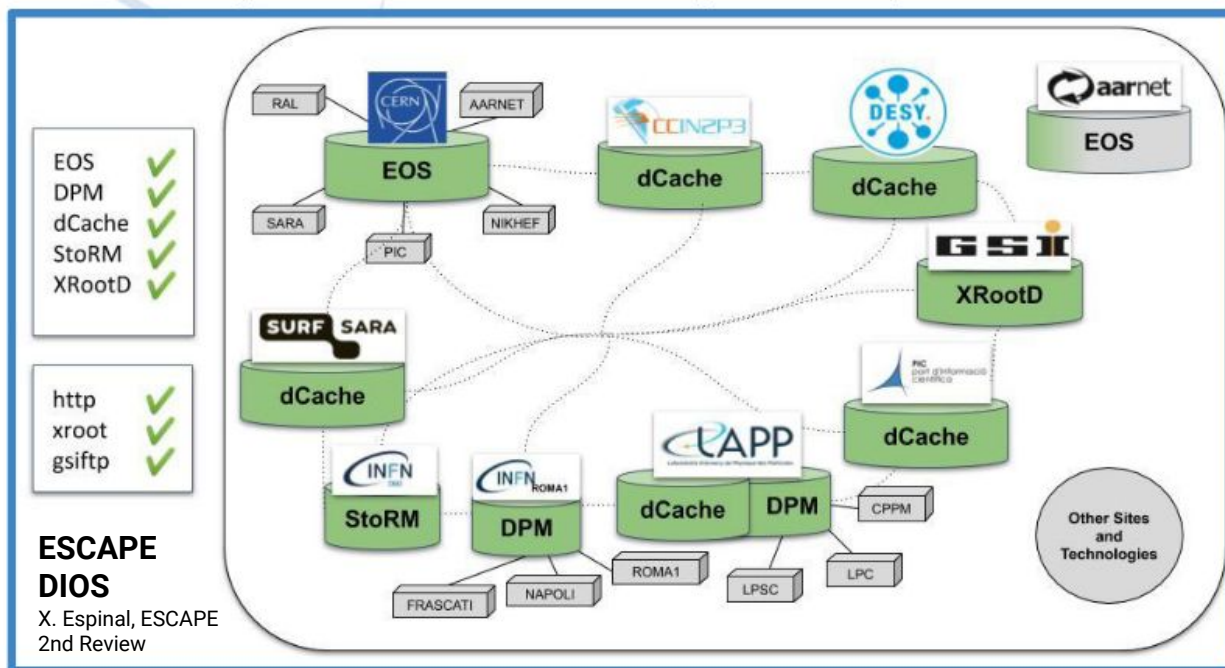
using HEP software for data management

Let's do science together
&
Thank you for your attention!