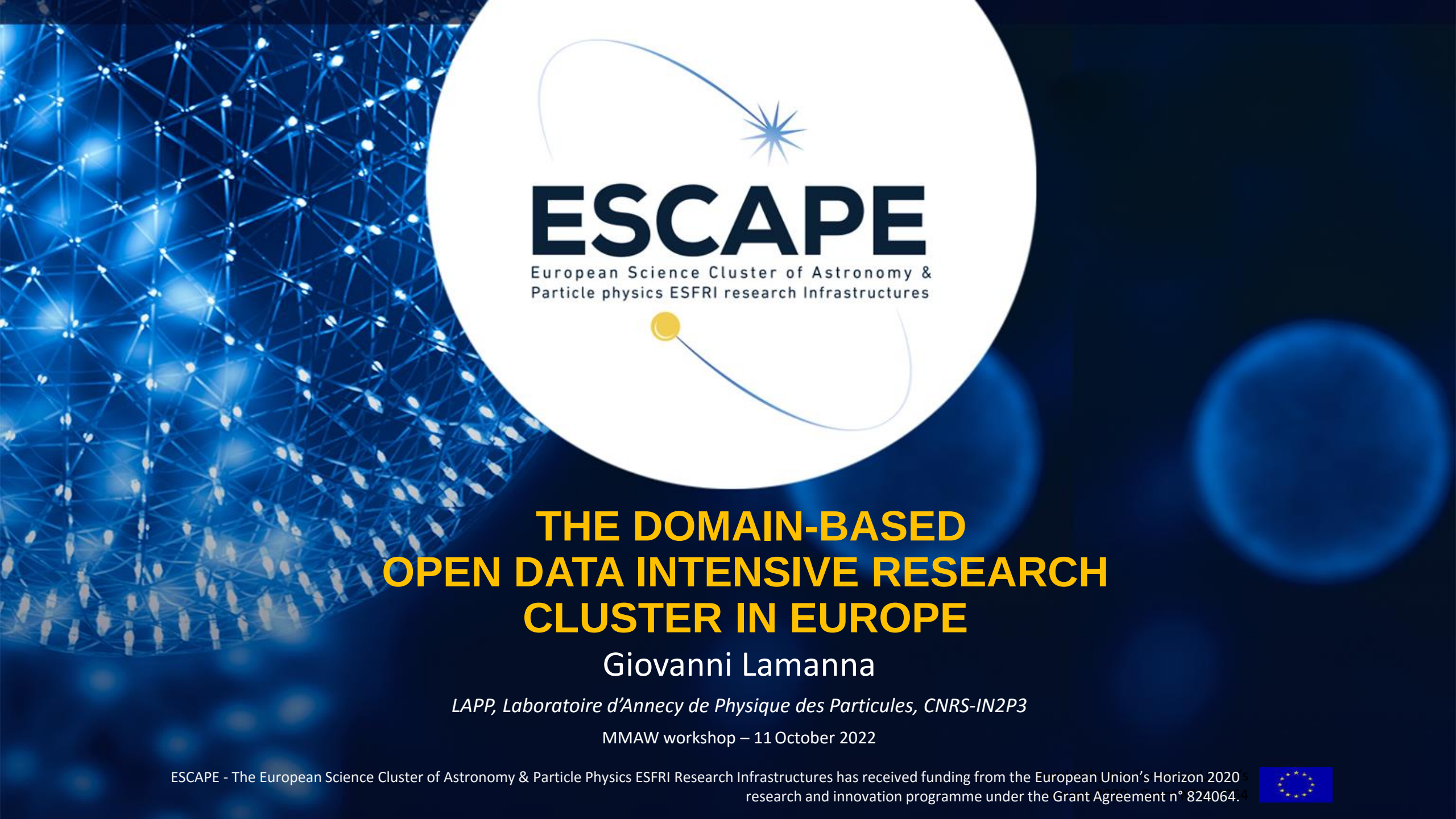


THE DOMAIN-BASED OPEN DATA INTENSIVE RESEARCH CLUSTER IN EUROPE

Giovanni Lamanna

LAPP, Laboratoire d'Annecy de Physique des Particules, CNRS-IN2P3

MMAW workshop – 11 October 2022

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

(ESFRI) and other RI PARTNERS

Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE H2020 CONSORTIUM

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

- 31 partners including 2 SMEs
- 10 ESFRI projects & landmarks: CTA, EST, FAIR, HL-LHC, KM3NeT, SKA, LSST, VIRGO, ESO, JIVE
- 2 pan-European International Organizations: CERN, ESO with their world-class established infrastructures, experiments and observatories
- 2 European Research Infrastructures: EGO and JIV-ERIC
- 1 involved initiative/infrastructure: EURO-VO
- 4 supporting European consortia: APPEC, ASTRONET, ECFA and NuPECC

- Budget: **15.98 M€**
- Started: **1/2/2019**
- Duration: **48 months** (end date 31/1/2023)
- Coordinator: **CNRS-LAPP**

Funded by the European Union's Horizon 2020 - Grant N° 824064

Five Science Clusters

ESCAPE is one of the five Science-Cluster projects that resulted from the H2020 topic call INFRAEOSC-04-2018: **"*Connecting ESFRI infrastructures through Cluster projects*",**
aimed at supporting:
"**Open data intensive driven science**" in order to "**rise productivity of researchers and to lead to new insights and innovation**"

Other Science Clusters: **ENVRI-FAIR** (Environment and Earth Sciences), **EOSC-LIFE** (Biomedical Science), **PANOSC** (Neutron and light sources facilities) and **SSHOC** (Social Science and Humanities).

Five Science Clusters

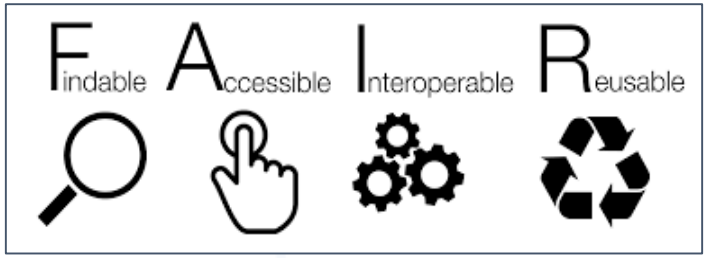
More than 80% of ESFRI RIs, plus other world-class RIs and new emerging ones.

The approach:

- Foster the establishment of **cross-border open innovation environment**.
- Develop **synergies and complementarity** between involved (ESFRI) research infrastructures.
- Adopt global standards and common solutions to the **data management for economies of scale**.

Working together making data and scientific research FAIR …

… as well as providing researchers with appropriate services and skills for data stewardship.



Link the ESFRI and other world-class RIs to **European Open Science Cloud (EOSC) …**

Funded by the European Union's Horizon 2020 - Grant N° 824064

European Open Science Cloud

EOSC in 2018:

A concept and a political decision of the EU member states in support of “Open Science”. According to the *EOSC Declaration :*

EOSC as data infrastructure commons serving the needs of scientists and federating resources.

Funded by the European Union’s Horizon 2020 - Grant N° 824064

ESCAPE: Astronomy, Nuclear and Particle Physics ESFRIs

Background analysis for the uptake of “Open Science” and “Data FAIRness”

- Builds on communities’ complementary excellences in data stewardship:
  - Astronomy Virtual Observatory infrastructure
  - HE-NP expertise in Exabyte-scale data management and large-scale distributed computing
- Builds on existing inter-RI synergies, intersections.
- Recognises that ESCAPE communities will be Exascale data generators, early adopters of ICT and data management innovations, push state-of-the-art.
- Both Observatory- and Facility- operations require global, open access to data, long term curation, and sustainability.
- Concerned by connecting the ESFRIs to EOSC - European Open Science Cloud (at the early stage)

Funded by the European Union’s Horizon 2020 - Grant N° 824064

ESCAPE work programme…
… aimed at building a domain-based «EOSC-cell»

Data Lake:
Build a scalable, federated, data infrastructure as the basis of open science for the ESFRI projects within ESCAPE.

Software Repository:
Repository of "scientific software" as a major component of the "data" to be curated in EOSC.

Science Platforms:
Flexible science platforms to enable the open data analysis tailored by and for each facility as well as a global one for transversal workflows.

Citizen Science:
Open gateway for citizen science on ESCAPE data archives and ESFRI community

Virtual Observatory:
Extend the VO *FAIR* standards, methods and to a broader scientific context; prepare the VO to interface the large data volumes of next facilities.

Funded by the European Union's Horizon 2020 - Grant N° 824064

Our domain-based "EOSC CELL"

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

ESCAPE OSSR

Catalogue & Repository of resources

Datasets
Software & services
Tutorials
Training
Publications

TSP's

RI-Specific Science Platforms

ESCAPE VO Virtual Observatory

Astronomy Data centres

VO Registry

VO Registry
Analysis Tools
VO Services

ESCAPE SAP Science Platforms

Workflows, notebooks, deployment platforms, packaging

ESCAPE CS Citizen Science

ESCAPE DIOS Data Lake

FAIR data management
Content discovery and delivery

HPC

PRACE

EuroHPC Joint Undertaking

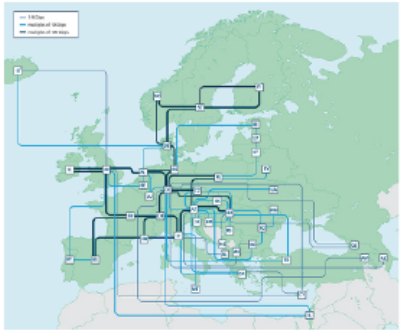
HTC

Grid clusters, etc

Private/public clouds

Commercial clouds

GÉANT

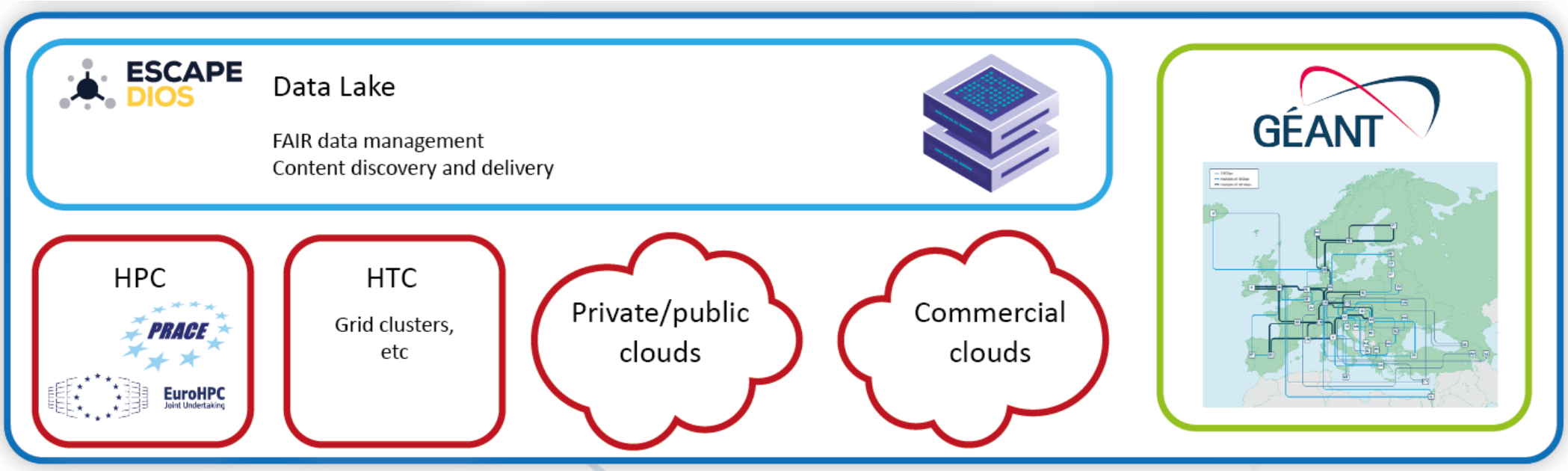


Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Infrastructure for Open Science (DIOS)

A functional Data Lake implemented (by all RIs and for their requirements)

- The ESCAPE pilot **Data Lake** (DL), after design and deployment was recently assessed, culminating in a joint exercise labelled as **Full Dress Rehearsal** (FDR20): a 24h production-like window where ESFRI's executed relevant workloads.
- Followed by a full scale **Data and Analysis Challenge** (DAC21) performing production-like Data Management, Processing and Analysis workloads including interplay possibilities using large scale resources (batch systems and clouds) and user-analysis oriented platforms (online notebooks and analysis platforms).



Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Infrastructure for Open Science (DIOS)

Full-cycle scientific data management and data processing

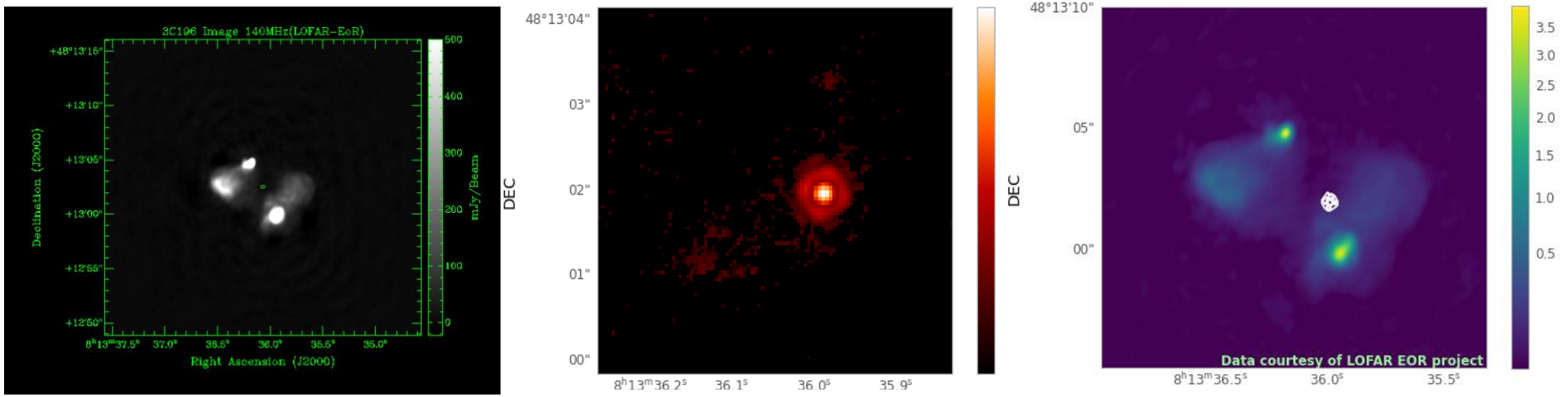
1. Data injected to the DL from **three** radio source observations in external locations

2. User in external location download the data, process and store results back to the DL

3. User interested in combining results stored with other public data to cover also visible spectrum

4. Combined optical data from the Hubble located via the **VO (WP4)**

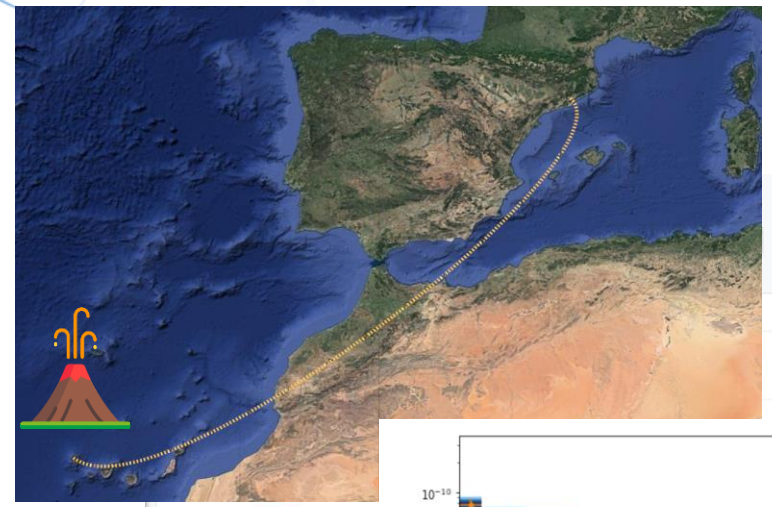
5. Optical and radio data aggregate in via the **ESAP (WP5), combined analysis done.** Results uploaded back to the DL.

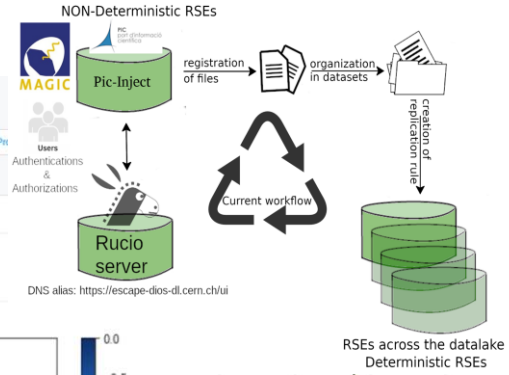


From left to right: Radio image, Optical image and the Combined image (LOFAR with optical contours)

Funded by the European Union's Horizon 2020 - Grant N° 824064

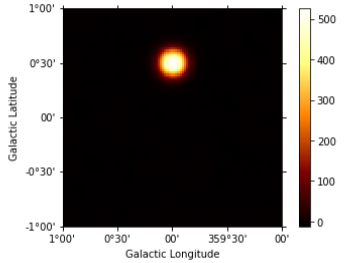
The ESCAPE Data Infrastructure for Open Science (DIOS)






```
[38]: # we can also compute the significance of our source
analysis.get_excess_map()
analysis.excess_map["sqrt_ts"].plot(add_cbar=True);
```

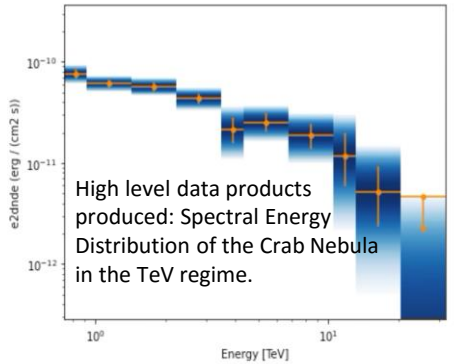
```
Computing excess maps.
Position <SkyCoord (Galactic): (l, b) in deg
    (0., 0.)> is outside valid IRF map range, using nearest IRF defined within
```



perform the fit

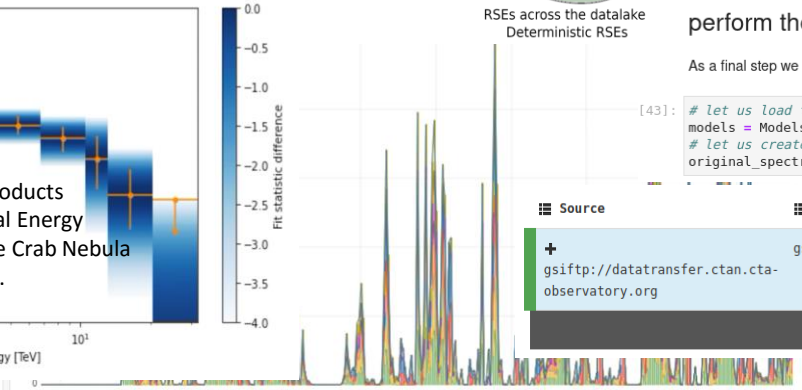
As a final step we fit the spectrum of the source, and we compare to the one we actually used for simulation

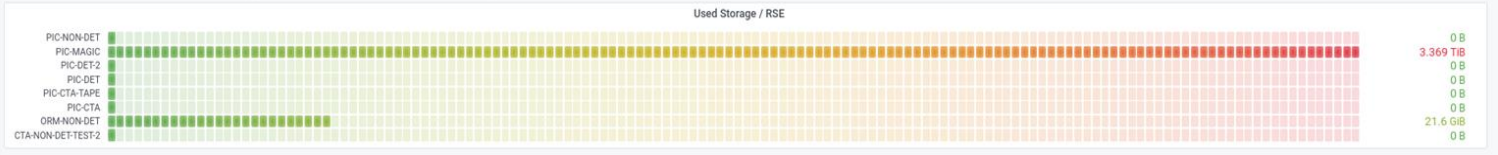
```
[43]: # let us load the model we used for the simulation
models = Models.read("./data/models/point-source-pwl.yaml")
# let us create a copy of the spectral model for later comparison
original_spectral_model = models[0].spectral_model.copy()
```



High level data products produced: Spectral Energy Distribution of the Crab Nebula in the TeV regime.

Source	Destination	VO	Submitted	Active
gsiftp://datatransfer.ctan.cta-observatory.org	gsiftp://door05.pic.es	pic01-rucio-server.pic	5589	129
			5589	129

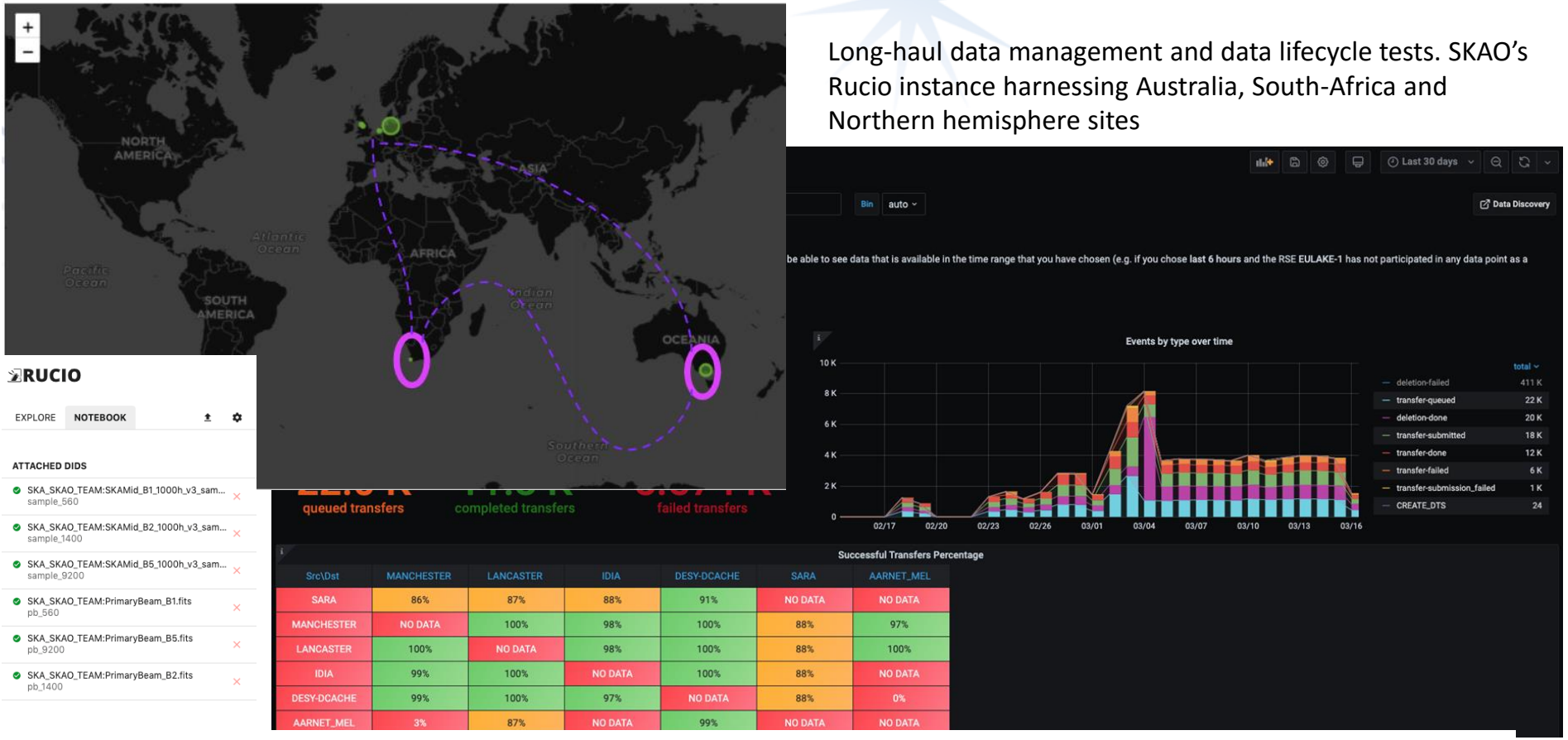




Funded by the European Union's Horizon 2020 - Grant N° 824064

The ESCAPE Data Infrastructure for Open Science (DIOS)

Long-haul data management and data lifecycle tests. SKAO's Rucio instance harnessing Australia, South-Africa and Northern hemisphere sites



Src\Dst	MANCHESTER	LANCASTER	IDIA	DESY-DCACHE	SARA	AARNET_MEL
SARA	86%	87%	88%	91%	NO DATA	NO DATA
MANCHESTER	NO DATA	100%	98%	100%	88%	97%
LANCASTER	100%	NO DATA	98%	100%	88%	100%
IDIA	99%	100%	NO DATA	100%	88%	NO DATA
DESY-DCACHE	99%	100%	97%	NO DATA	88%	0%
AARNET_MEL	3%	87%	NO DATA	99%	NO DATA	NO DATA

Manual data transfers with Rucio began Feb 18th 2021 and automated tests running hourly across all sites since Feb 23rd

Funded by the European Union's Horizon 2020 - Grant N° 824064

Overview progress: implementing the EOSC-cell

Deployment of the Open-source ESCAPE catalogue of resources (data and software)

Shared domain-based open science software and services based on FAIR principles

ESCAPE OSSR

Catalogue & Repository of resources

Datasets
Software & services
Tutorials
Training
Publications

Objectives:

- Facilitate and support continuous **development, deployment, exposure and preservation** of partners' software/tools/services
- Foster **interoperability, software re-use and cross-fertilisation** between ESFRIs
- Offer an **open innovation environment for open standards** (e.g. workflows, data-formats), **common regulations** and **shared (novel) software** for **multi-messenger & multi-probe data**

eOSSR

DOI 10.5281/zenodo.5524912

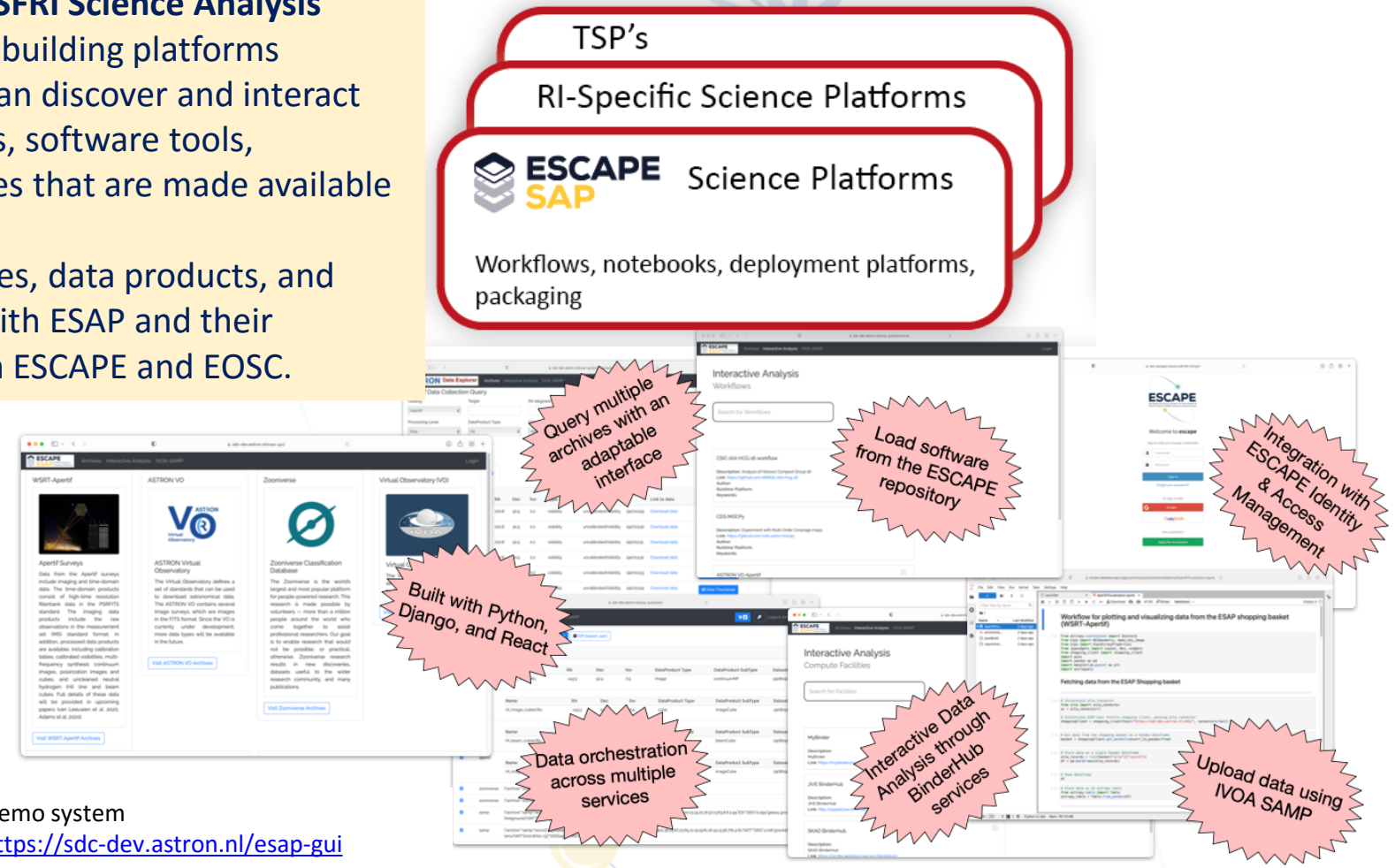
eOSSR library https://gitlab.in2p3.fr/escape2020/wp3/eossr

- *incorporates all OSSR developments, based on the commonly defined practices and standards*
- *python-based*
- *OSSR API : send request to the OSSR, find and filter software and services, upload new entries, update existing entries*
- *CI : automated upload / update using gitlab CI*
- *Metadata : schema definition, crosswalk between CodeMeta and Zenodo, generator and validator available*

Funded by the European Union's Horizon 2020 - Grant N° 824064

Overview progress: implementing the EOSC-cell

- Development of the **ESFRI Science Analysis Platform**: a toolkit for building platforms through which users can discover and interact with the data products, software tools, workflows, and services that are made available through ESCAPE.
- Preparing ESFRI services, data products, and tools for integration with ESAP and their subsequent use within ESCAPE and EOSC.

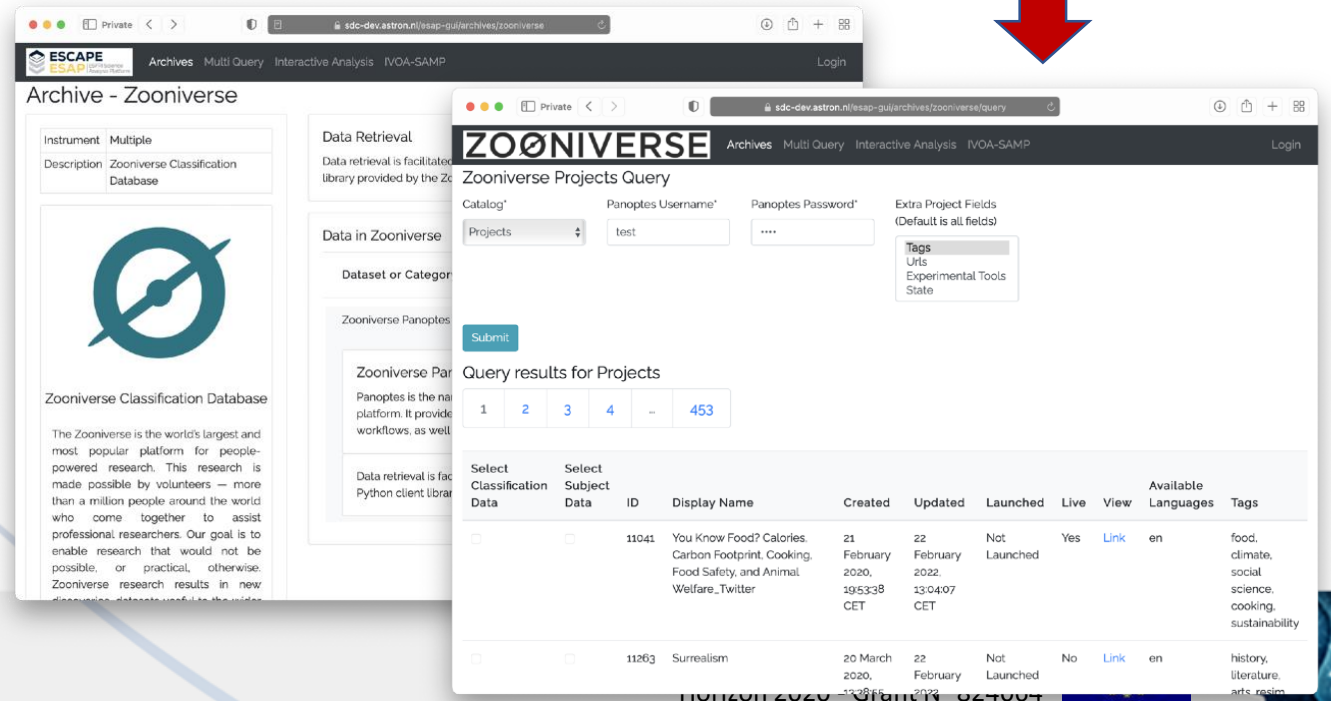


Demo system
https://sdc-dev.astron.nl/esap-gui

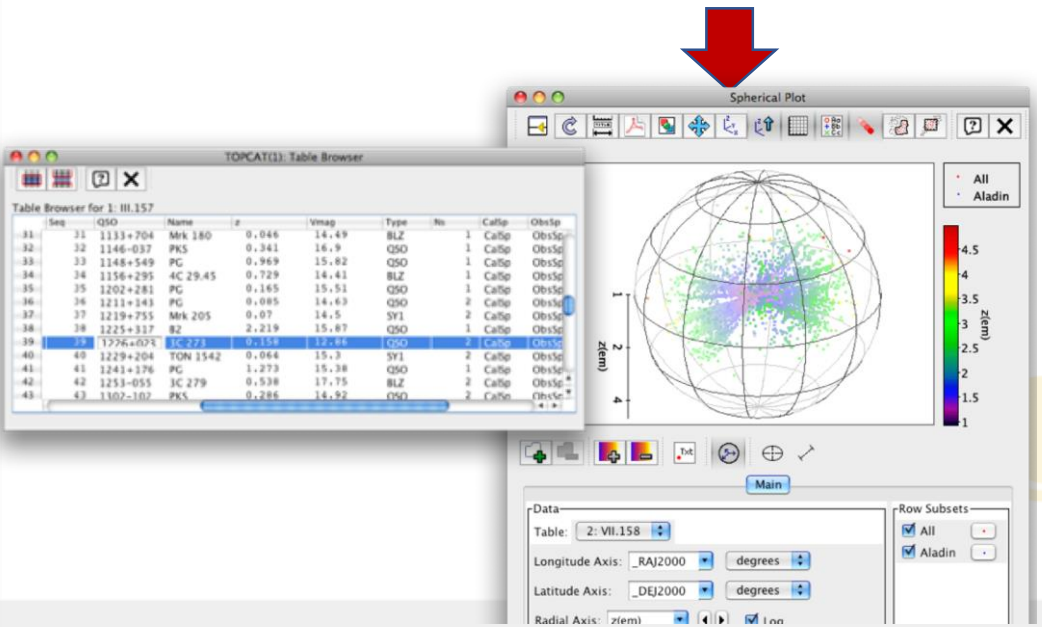
Funded by the European Union's Horizon 2020 - Grant N° 824064

Overview progress: implementing the EOSC-cell

Full integration of ESAP with the ESCAPE Zooniverse "Panoptes" system.



Overview progress: implementing the EOSC-cell



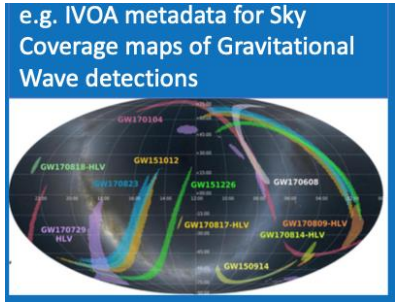
Enables a *Virtual Research Environment* of interoperable tools and services based on IVOA standards, including ESO Science Archive and ESO Science Portal

Funded by the European Union's Horizon 2020 - Grant N° 824064

Overview progress: implementing the EOSC-cell

Multi-Order Coverage 2.0
Space coverage extended with TIME coverage.

Driven by ESFRI/RI needs (EGO, ESO, Radio astronomy,+++).



ESFRI / RIs	Results for ESCAPE work toward FAIR standards and tools
ESO-ELT	- Data access and visualisation standards and tools - Support of **VO standards in ESO archive** services – used as exemplary case to help others - Relevant IVOA standards updated
EGO/VIRGO	- Development of **MOC2.0 and mocpy** - Tools / libraries integrated into GW community software - Paper published in Astronomy & Computing
SKA, JIVE, ALMA (LOFAR)	- Creation and support of the **IVOA Radio Astronomy** Interest Group - Example TAP services, accessible in VO tools and in the ESCAPE platform
CTA & KM3NeT	- Data Provenance standards approved by IVOA - Many activities for adoption and implementation (Workshop held) - Reference paper published on a: ***Management System for Provenance Information***
EST	- VO metadata developed for Solar Physics - Prototype TAP services for solar data

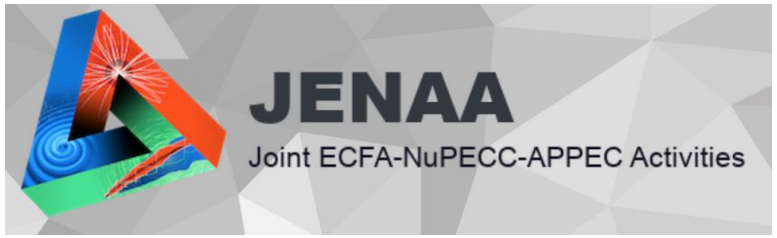
Funded by the European Union's Horizon 2020 - Grant N° 824064

Enhancing participation in EOSC

Test Science Projects – TSPs

- Originally proposed to validate ESCAPE services for Open Science.
- Proposed to demonstrate **multi-messengers / multi-domain** science integration across ESCAPE
- As a “bench” concept aiming at enhancing researchers participation in open science and cross-domain scientific research.
- Benefit real science goals in exploring synergies between the ESFRIs and largely among three scientific communities Astrophysics/Astroparticle, accelerator-based Particle and Nuclear Physics (supported by Joint ECFA, NuPPEC, APPEC Activities (JENAA).

1. Dark Matter Science Project
2. Extreme Universe (& Gravitational waves) Science Project



ESCAPE implements a Virtual Research Environment (VRE) hosting the pilots of the two TSPs:
thematic collaborative digital environment used by scientists, which enables FAIR community-based scientific research, training, innovation, cross-fertilisation and open science.

Funded by the European Union’s Horizon 2020 - Grant N° 824064

Extreme Universe Science Project

A series of pilots focused on violent phenomena in the Universe with **Astrophyiscal** as well as **fundamental implications**

Collecting requirements for VRE.

Understanding **services**, **computing resource needs and technical challenges.**

First full data analysis results expected for October 2022 during the "**ESCAPE to the future**" workshop

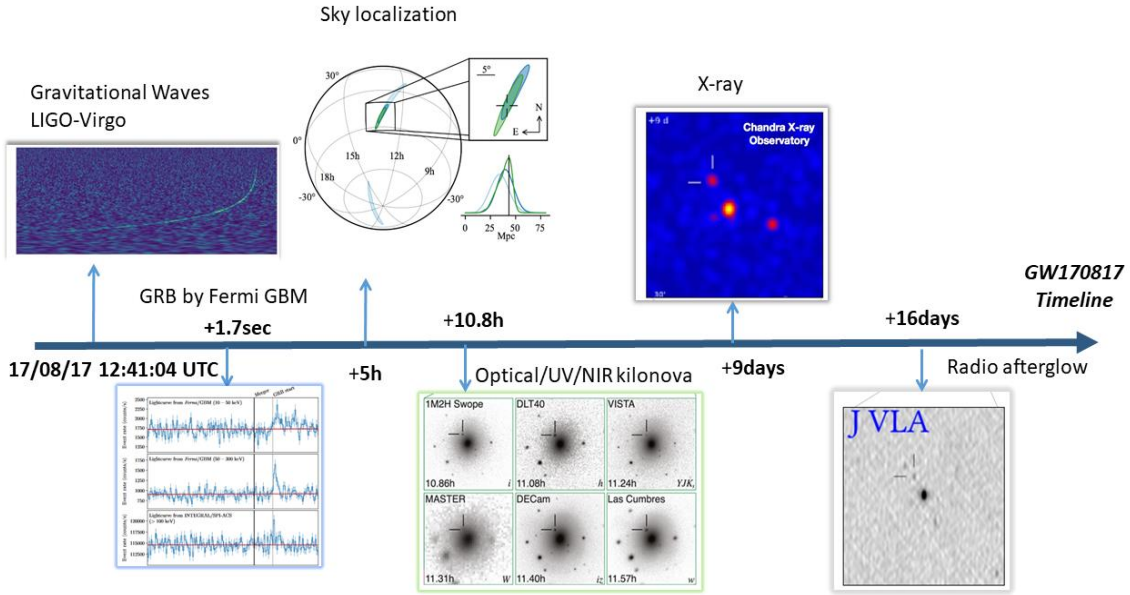
Main Research Area	Objects/sources	Messengers	ESF/RI involved	ESCAPE services EOSC-Future integrations	Data Analysis tools (AI,ML)	Pilot project(s)	Computing resources required
Compact objects	*Pulsars, FRBs, Off-nuclear AGN*	*radio, optical, X-ray, …*	*LOFAR...*	*Multiwavelength platform/Software catalogue,VO tools*	*Data science, Machine Learning*	*1) Radio astronomy: FRBs, pulsars, plerions, off-nuclear AGN*	*Compute cluster, Jupyter hub, Rucio Data lake*
High energy Astrophysics	*GRBs, jets, AGN, BNS, CCSN*	*neutrinos, gamma-ray, radio,X-ray, GW,...*	*CTA, Virgo, KM3NeT, SKA,LSST*	*Multimessenger platform/Software catalogue,... Virtual Observatory tools*	*Model comparison, Machine Learning*	*1)GRB/neutrino/GW analysis, 2) Blazar MWL/neutrino*	*GPU cluster Jupyter hub*
Fundamental physics	*Dark matter, GR, Primordial Universe*	*GW,*	*Virgo, Einstein Telescope*	*Template banks, generation software,...*	*Machine learning approach*	*1) DM template bank and ML analysis pipeline*	*GPU cluster Jupyter hub*

Funded by the European Union's Horizon 2020 - Grant N° 824064

Extreme Universe Science Project

Multi-Messenger Astrophysics

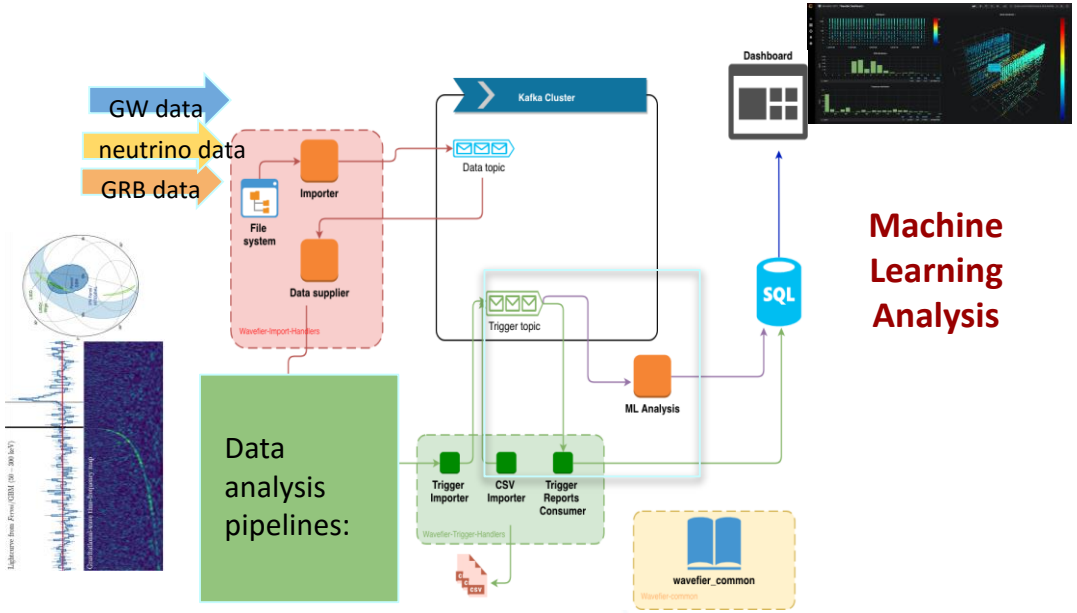
- **Analyzing Multimessenger signals within ESCAPE/EOSC framework with a real time transient signal classifier**



Abbott et al. 2017

Wavefier: A prototype for Real time analysis

- *Multimessenger platform/Software catalogue*
- *GRB/neutrino/GW analysis*
- *EOSC implementation of real time classifier*



Machine Learning Analysis

Elena Cuoco (EGO)

Funded by the European Union's Horizon 2020 - Grant N° 824064

European Open Science Cloud

EOSC in 2018:

A concept and a political decision of the EU member states in support of "Open Science". According to the *EOSC Declaration :*

EOSC as data infrastructure commons serving the needs of scientists and federating resources.

ESCAPE (and the SCL) having a major scientific role along such an evolution

EOSC in 2022:

« EOSC Association » legal entity established with more than 200 members, steering investments via its Task Forces and other governance structures.

Advancing Open Science to accelerate the creation of new knowledge, inspire education, spur innovation and promote accessibility and transparency.

With the ambition to develop a « **Web of FAIR Data and services** » for science in Europe (as achieved through the VRE of ESCAPE).

Funded by the European Union's Horizon 2020 - Grant N° 824064

The new remit of the Science Clusters in Horizon Europe

2023-2027:

- **Consolidate common EOSC approaches between the RI communities** involved in the five science clusters and extend outreach towards new or underrepresented user communities.
- Firmly install within the EOSC ecosystem the implementation of open science practices and the management of FAIR research digital objects into the core operation of ESFRI and other RIs.
- Test **financial models for cross-RI service** provision through the EOSC.
- Establish a mechanism to **collect operational needs coming from the user communities and to interact with future operator(s) of the EOSC platform.**
- Demonstrate and pilot the use of EOSC resources by multiple research communities through cross-RI and/or cross-domain **open science projects** and services

Funded by the European Union's Horizon 2020 - Grant N° 824064

New ESCAPE Collaboration Agreement

Science Clusters as domain-based "coordinating structures" for Open Science and operationalising the European Research Area (ERA).

ESCAPE after the H2020 grant becomes a sustained "Community Platform" established by a **Collaboration Agreement** to operate Open Science as well as cooperating in order to address new topics in ERA.

https://indico.in2p3.fr/event/24327/

Funded by the European Union's Horizon 2020 - Grant N° 824064

New ESCAPE Collaboration Agreement

ESCAPE collaboration preliminary Work Plan:

For EOSC destination and

- Data Lake
- Software catalogue, methods and training
- European Virtual Software Institute for Research
- Virtual Research Environments
- Open Science Research Projects
- Data interoperability

European Research Area (ERA) operational

- Low Latency data analysis
- Technological dev. and innovation (e.g. quantum)
- Curiosity driven RIs' synergies
- Inter-cluster projects for society
- Citizen Science
- Researchers' rewarding
- Sector Data Spaces […]

Funded by the European Union's Horizon 2020 - Grant N° 824064

About HORIZON-INFRA-2023-SERV-01-02

Research infrastructures services advancing frontier knowledge.

- "Proposals should make available to researchers a very wide and comprehensive portfolio of complementary research infrastructures services, **including data services**, which are relevant for frontier research in the domain."
- **"The results as the main objective of these actions is the service provision** (applicants are not required to include in their proposal a plan for the exploitation and dissemination)."

However:

- "Data management, interoperability, as well as the connection of digital/data services to the **European Open Science Cloud**, should be addressed where relevant."
- "Proposals should duly take into account major **European or international initiatives** relevant in the domain. Whenever appropriate, they **should foster the use and deployment** of (open) global standards."
- "Project results are expected to contribute to a better management, **including implementing FAIR data principle**, of the continuous flow of data collected or produced by research infrastructures."

Therefore:

- Biosphere: terrestrial biodiversity and ecosystems, including forest; ⇒ ENVRI-FAIR cluster
- **Astronomy and Astroparticle physics;** ⇒ **ESCAPE cluster**
- Arts and Humanities. ⇒ SSHOC cluster

Funded by the European Union's Horizon 2020 - Grant N° 824064

About HORIZON-INFRA-2023-SERV-01-02

ESCAPE services to be exploited

- ✓ Data Lake
- ✓ Software catalogue, methods and training
- European Virtual Software Institute for Research
- ✓ Virtual Research Environments
- Open Science Research Projects
- ✓ Data interoperability

Common interests

- ★ Low Latency data analysis
- Technological dev. and innovation (e.g. quantum)
- ★ Curiosity driven RIs' synergies
- Inter-cluster projects for society
- ★ Citizen Science
- Researchers' rewarding
- Sector Data Spaces […]

Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

https://projectescape.eu/

https://projectescape.eu/events/escape-future

THANK YOU!

Funded by the European Union's Horizon 2020 - Grant N° 824064