The ESCAPE Cluster for Open Science in the EOSC

Cristiano Bozza - University of Salerno and INFN

COST CA18108 Workshop on future challenges and opportunities in QGMM - Napoli (Italy)

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.

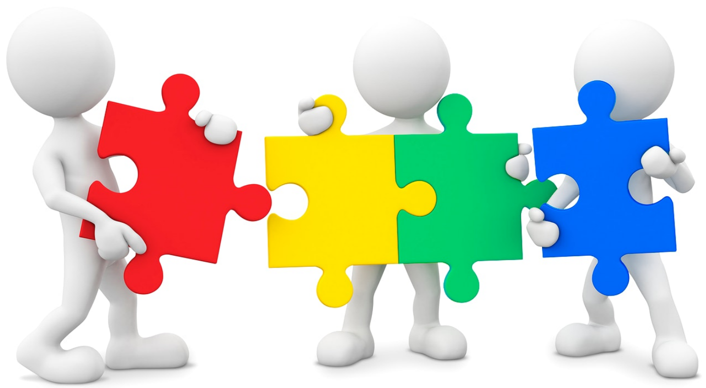
The FAIR approach and Open Science

All latest breakthroughs in science are the result of expensive, long-lasting, personpower-intensive efforts

AND

The scientific method is all about reproducibility of experiments, analyses and results!

- **Share**
- **Cooperate**
- **Cross-check**
- **Cross-fertilize**



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

The FAIR approach and Open Science

The FAIR approach (see https://www.go-fair.org/fair-principles/):

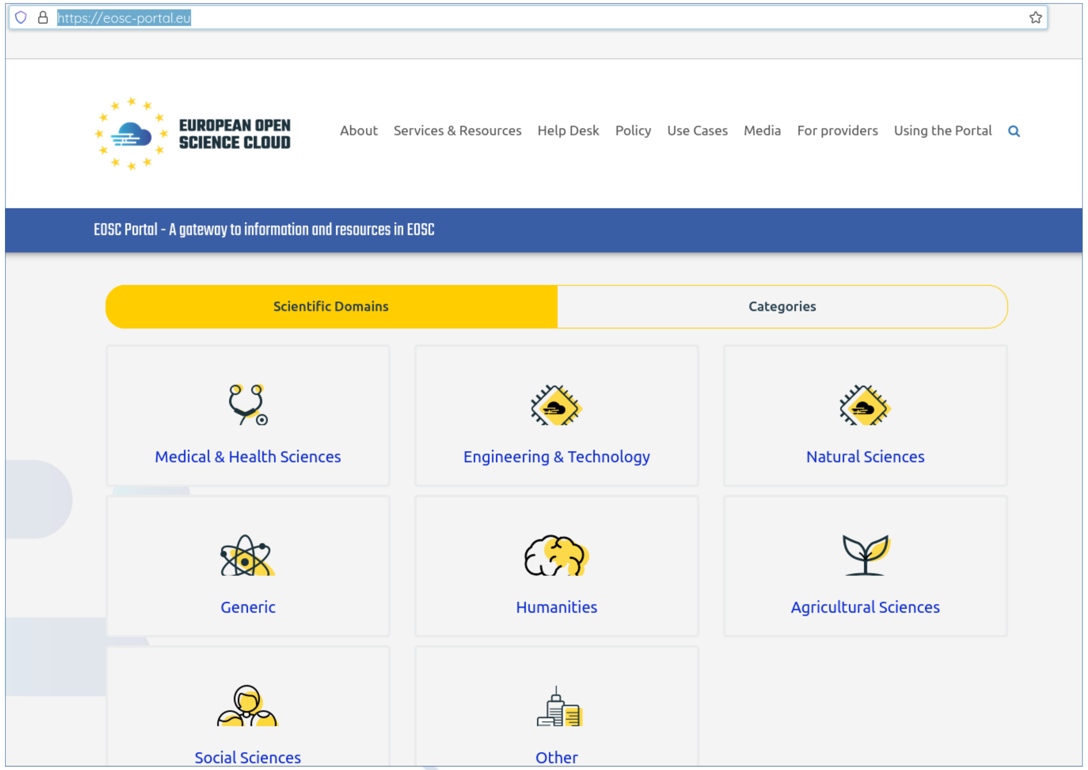
- **F**indable: *make data automatically findable, by means of metadata that are easy to understand both to humans and machines*
- **A**ccessible: *once found, data should be easy to access, both for human and machines, with proper authorization/authentication*
- **I**nteroperable: *data should be stored and retrieved in ways that allow them to be integrated with other data and used in different analysis workflows*
- **R**eusable: *metadata and data should be well-described so that they can be reproduced or combined in different settings*

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

FAIR data and Open Science: the EOSC

- The European Open Science Cloud aims “to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes.” (from the EC website)

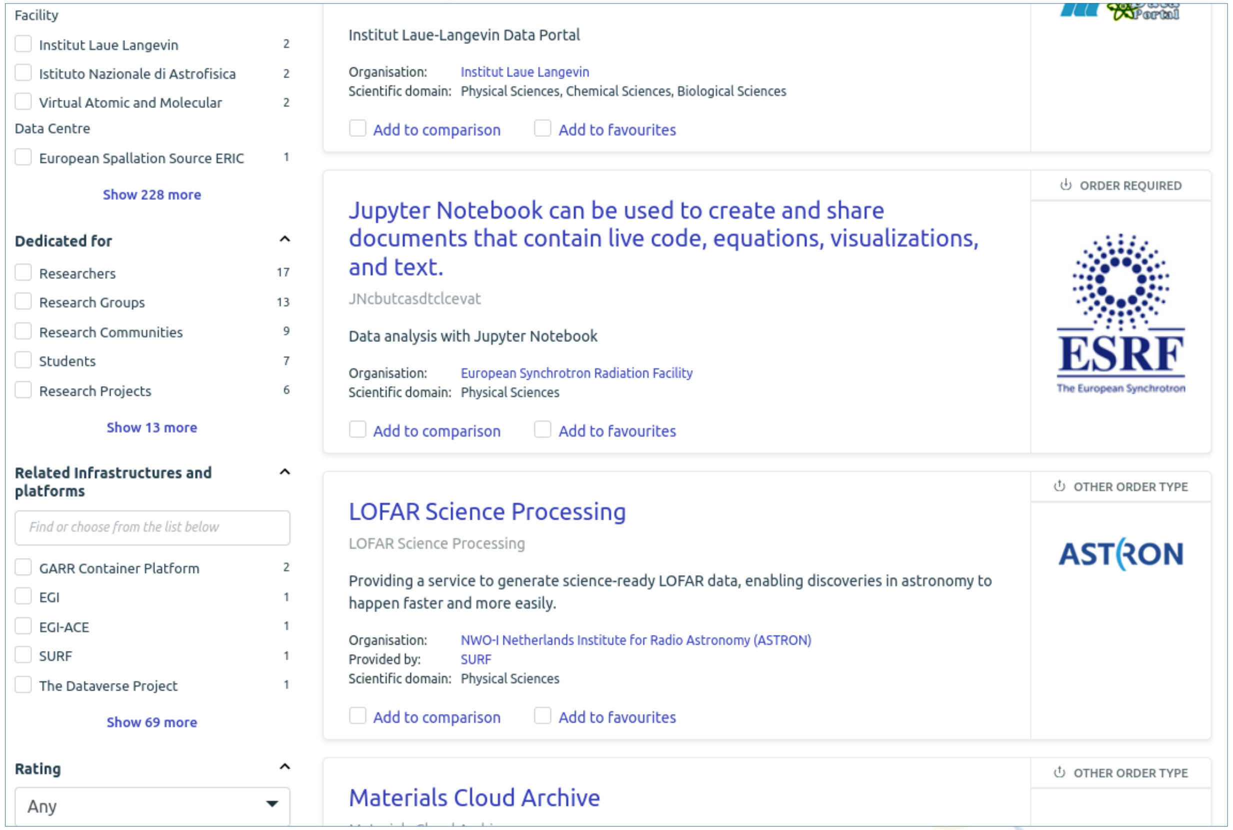
- Access portal: https://eosc-portal.eu/

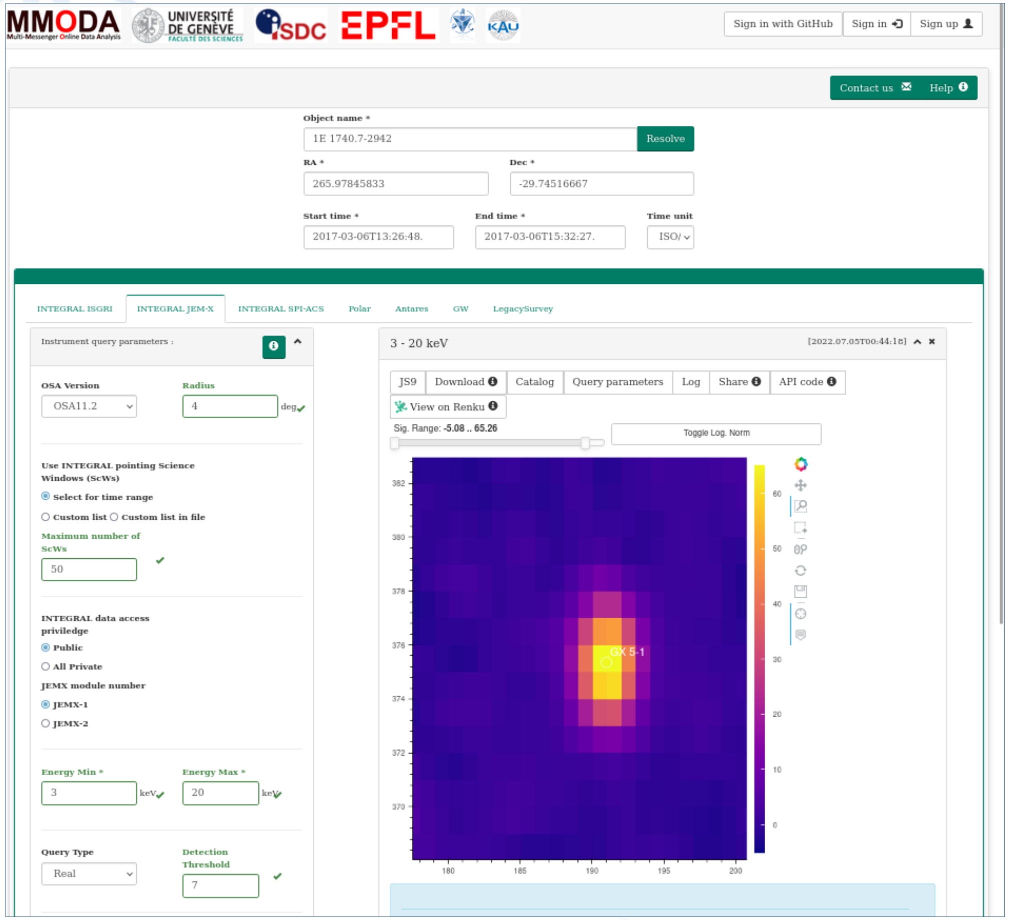


Funded by the European Union’s Horizon 2020 - Grant N° 824064

FAIR data and Open Science: the EOSC

- Marketplace and catalogue in the portal (https://eosc-portal.eu/)





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

ESCAPE overview

https://www.projectescape.eu/

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

ESCAPE overview

ESCAPE partners: Astrophysics & Particle Physics Research Infrastructures

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

ESCAPE overview

- Development/activity areas at a glance

TSP's

RI-Specific Science Platforms

ESCAPE SAP Science Platforms

Workflows, notebooks, deployment platforms, packaging

ESCAPE OSSR

Catalogue & Repository of resources

Datasets
Software & services
Tutorials
Training
Publications

ESCAPE VO Virtual Observatory

Astronomy Data centres

VO Registry

VO Registry
Analysis Tools
VO Services

ESCAPE CS Citizen Science

ESCAPE DIOS Data Lake

FAIR data management
Content discovery and delivery

HPC

PRACE

EuroHPC Joint Undertaking

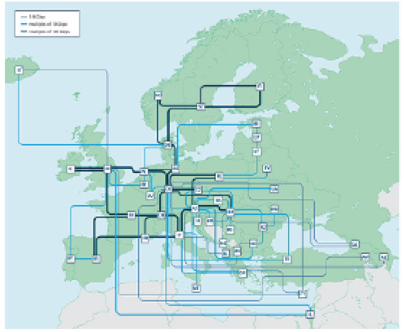
HTC

Grid clusters, etc

Private/public clouds

Commercial clouds

GÉANT



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

ESCAPE DIOS

FAIR data management

DIOS: a data-lake infrastructure in ESCAPE

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

ESCAPE DIOS

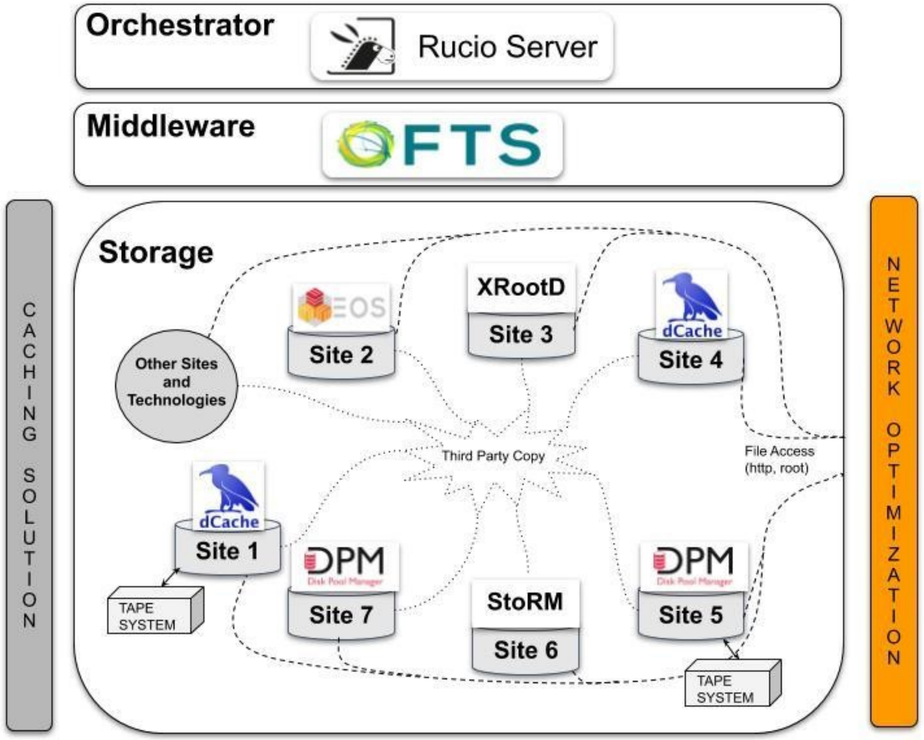
DIOS: a “data lake” prototype infrastructure to support FAIRness of data at Exabyte scale

- Storage services
  - PB-scale capacity available deployed at DESY, SURF-SARA, IN2P3-CC, CERN, IFAE-PIC, LAPP, GSI and INFN(CNAF, ROMA and Napoli)
- Data management and orchestration service: RUCIO (from ATLAS/LHC)
  - Instances running now at CERN, SKAO, IN2P3 and PIC
- File transfer and data movement: FTS (from WLCG/CERN)
  - Instances at CERN and PIC and using various protocols: http, gridftp and xrootd (and swift for HPCs)
- Content delivery and latency hiding: XCache (from XrootD/SLAC)
  - Deployments at GSI, CERN, LAPP and IN2P3
- Global Information System: CRIC (from ATLAS/LHC) deployed at CERN
- Auth/Authz/IM (AAI): Indigo IAM (from INDIGO-DATACLOUD, AARC/AARC2 and FIM4R projects). The ESCAPE IAM instance is deployed and operated in INFN/CNAF

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

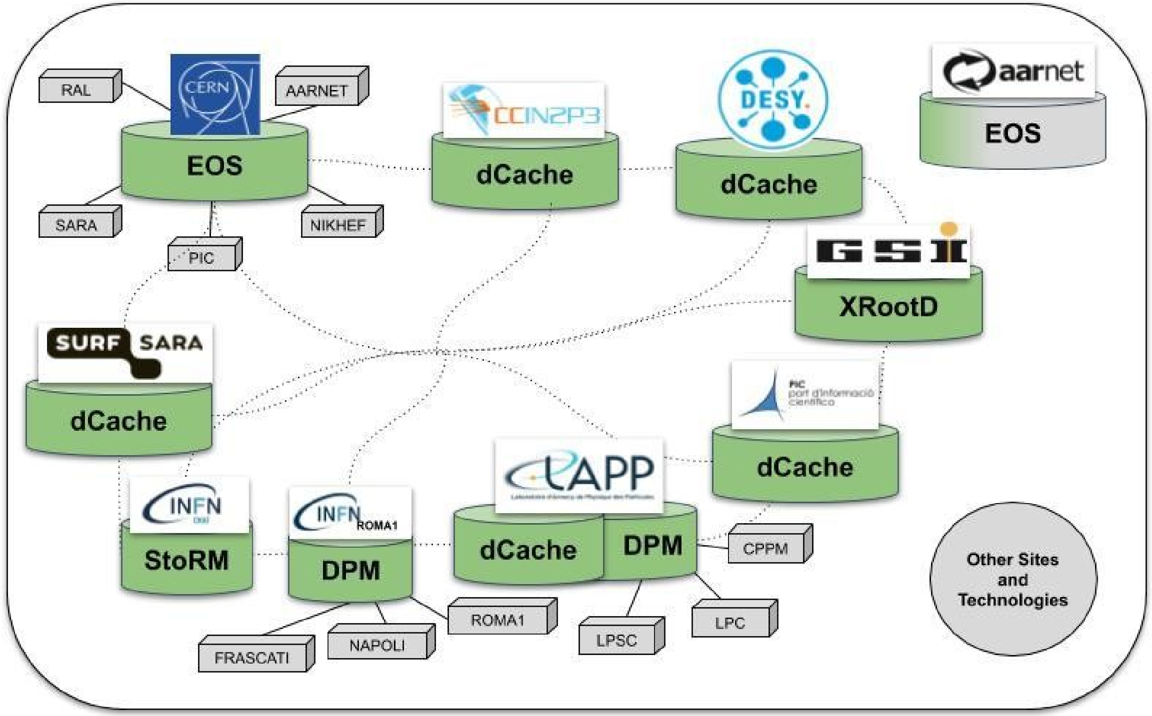
ESCAPE DIOS

- DIOS Resource allocation



EOS ✓
DPM ✓
dCache ✓
StoRM ✓
XRootD ✓

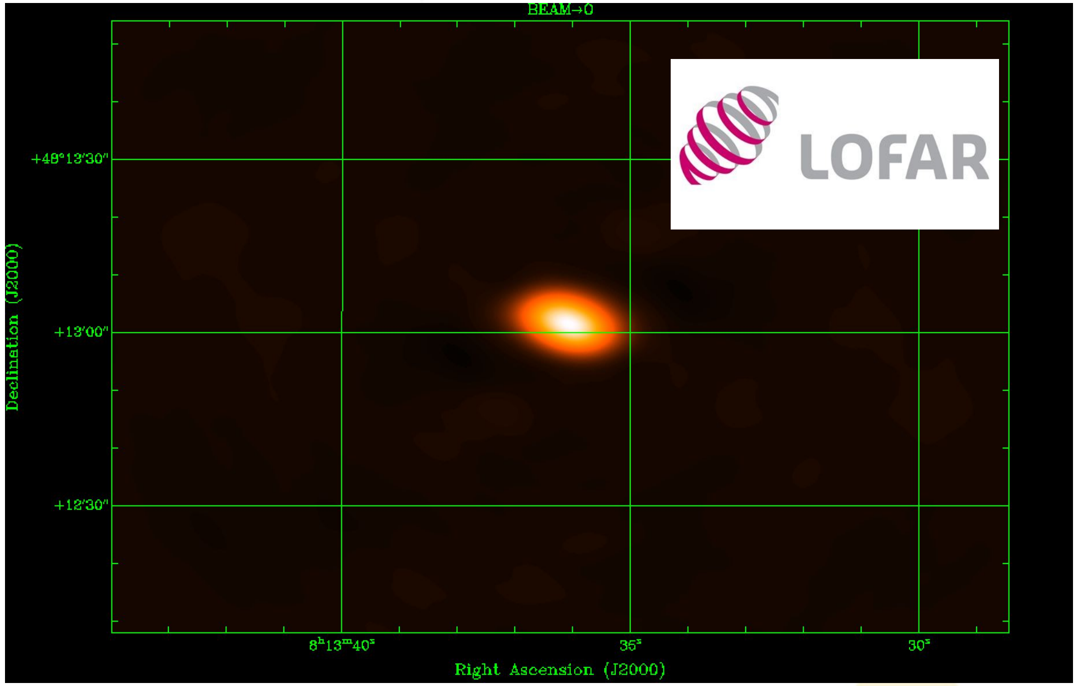
http ✓
xroot ✓
gsiftp ✓



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

ESCAPE DIOS

- DIOS usage example



LOFAR: astronomical radio source 3C196 made using LOFAR data. The raw visibility data was downloaded via RUCIO from the EULAKE-1 and processed on Open Nebula at SurfSara using the container based LOFAR software

EGO/VIRGO: Upload 4h of Virgo public data sampled at 4kHz from an EGO server to the datalake. Download them to CNAF. The data is split into 1s samples. Making available the real-time strain data to pipelines and tools assessing the data quality.

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

ESCAPE DIOS

- DIOS usage example

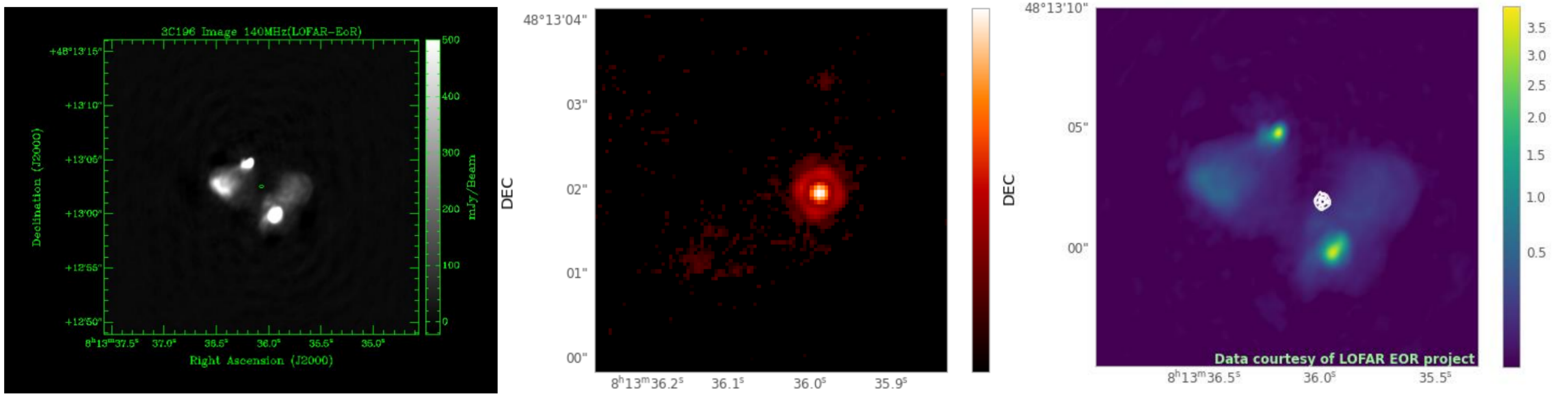
1. Data injected to the DL from **three** radio source observations in external locations

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

ESCAPE OSSR

FAIR and Open software

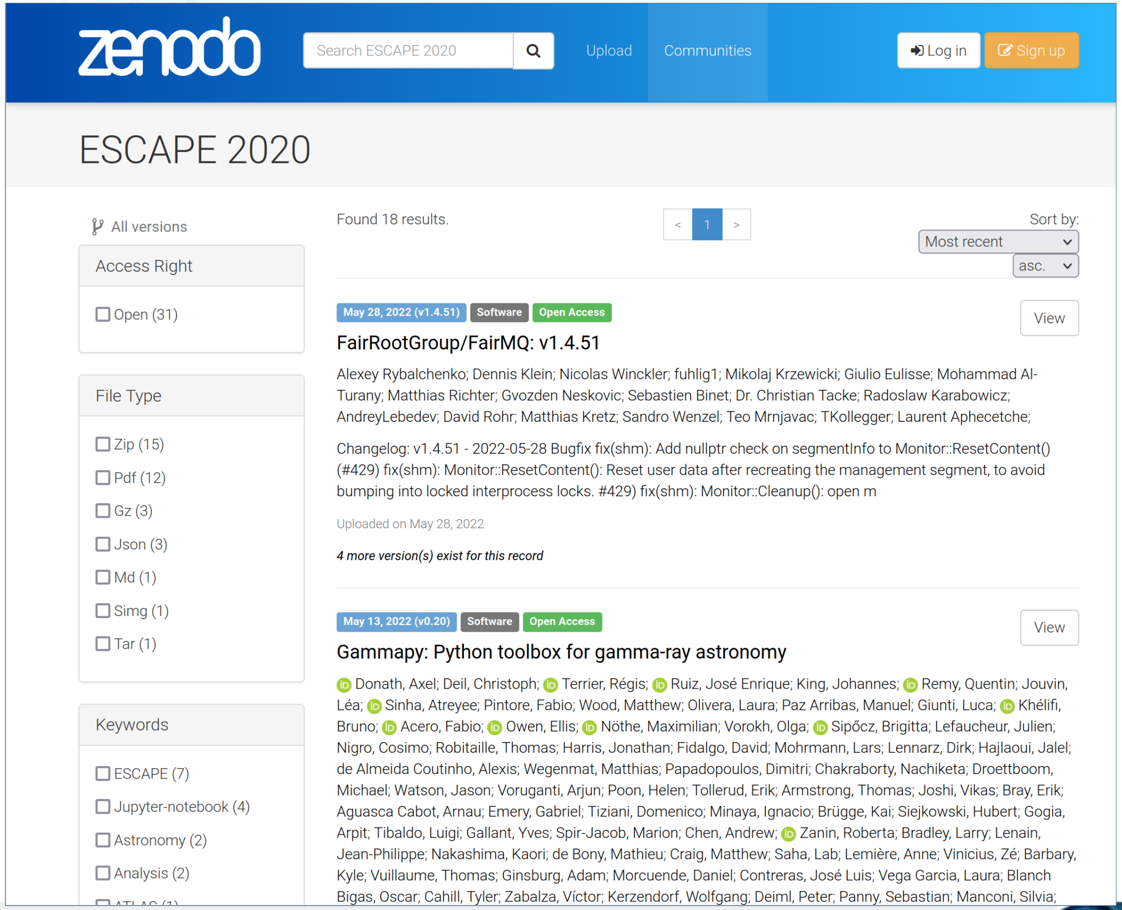
OSSR: Open Source Scientific software Repository

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

ESCAPE OSSR

- The goal: populate the EOSC with a (sustainably) growing library of open source software for scientific research

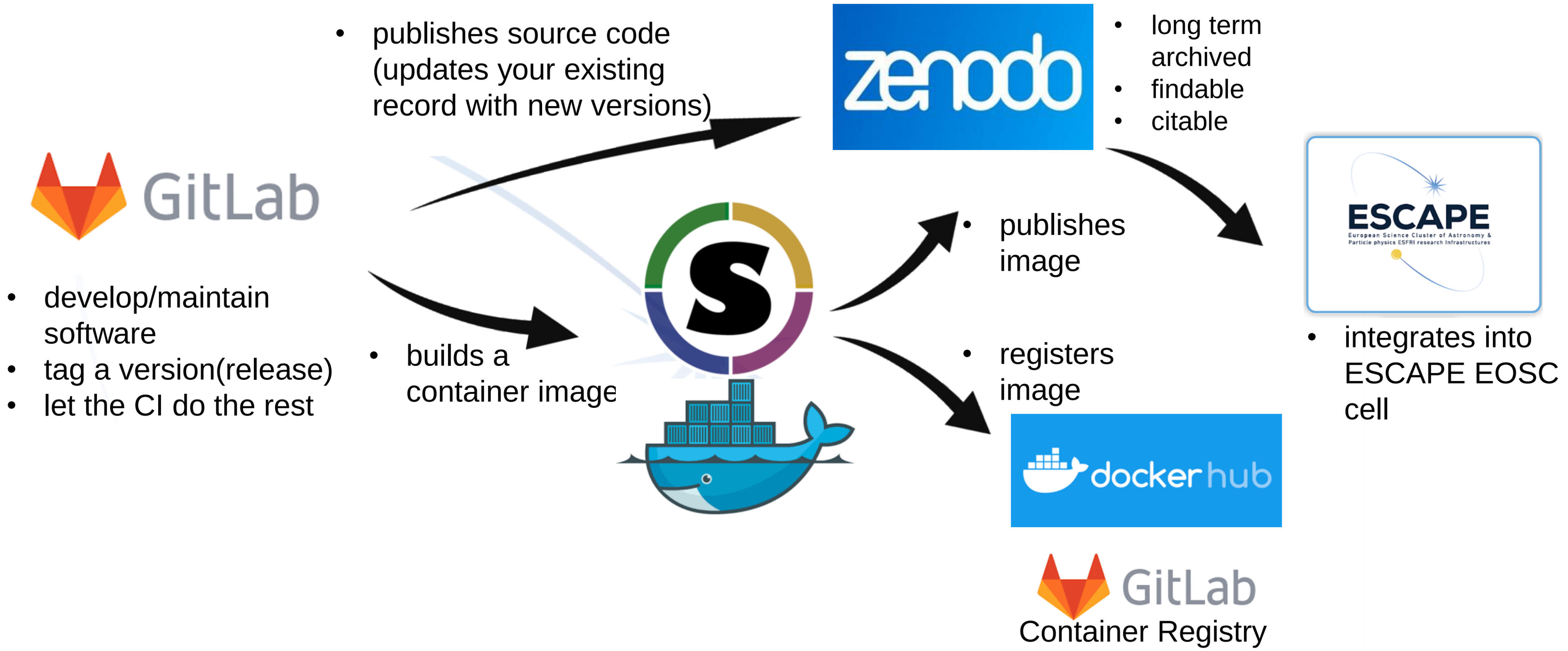
https://escape2020.pages.in2p3.fr/wp3/ossr-pages/



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

ESCAPE OSSR

- OSSR development platform: not only repository but also a working tool



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

ESCAPE OSSR

- OSSR Software showcase (incomplete!)

Gammalearn

HCG-16 study

km3py

LOFAR SW

agnpy

gLike

gammapy

ConCORDIA

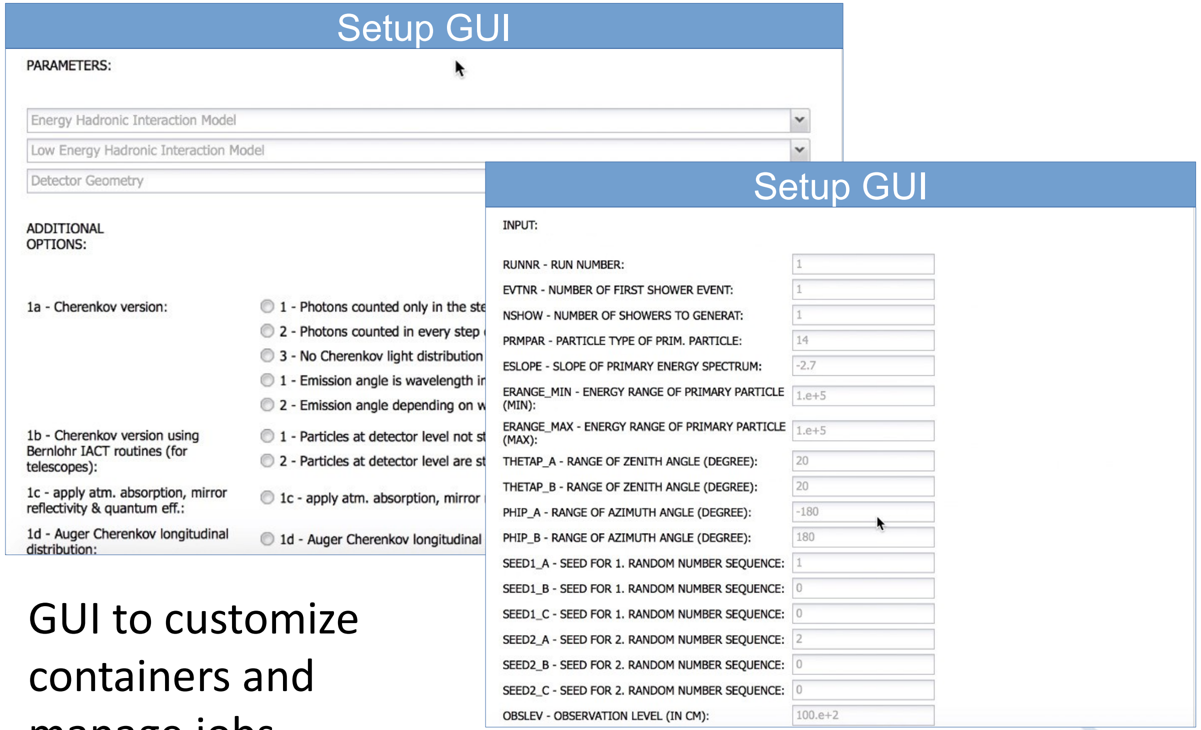
FairRoot & Fair SW

SKA data challenge

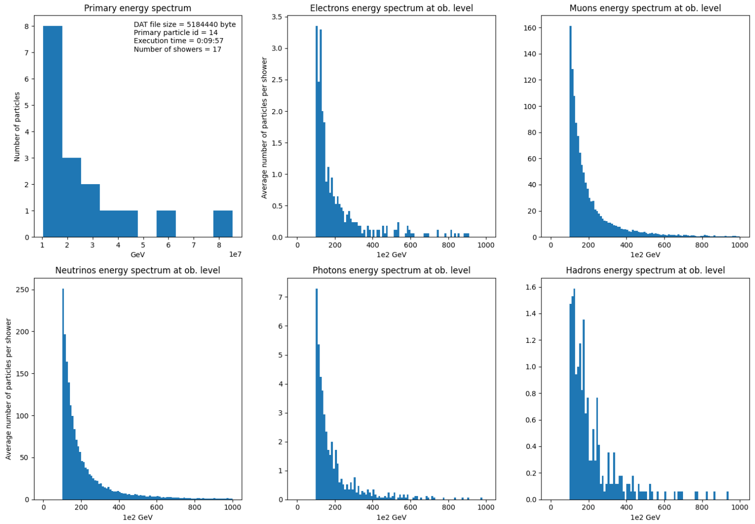
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE OSSR

- OSSR Example: ConCORDIA – **Con**tainers for **COR**SIKA on **DI**RAC
  - A set of turnkey containers for cosmic ray simulation



GUI to customize containers and manage jobs



Performance footprint published together with container

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

ESCAPE OSSR

- OSSR Training
  - Software as first class EOSC citizen
  - Enable software custodians

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

ESCAPE CEVO

Connecting ESFRI projects to EOSC through the Virtual Observatory

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

ESCAPE CEVO

- Aim - make Astronomy data and services FAIR to support Open Science

- Objectives :
  - Assess and implement the connection of ESFRI and other astronomy research infrastructures to the EOSC by the Virtual Observatory.
  - Refine and pursue implementation of FAIR principles for astronomy data via common interoperability standards - extending the VO to new communities.
  - Establish data stewardship practices for adding value to scientific content of ESFRI data archives.

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

ESCAPE CEVO

- The Virtual Observatory is an operational framework for interoperable access to astronomical data and services
- A pioneer of FAIR data sharing - an existing global framework – populated by major data providers (space and ground based) that is heavily used by the community (e.g. ESA Gaia mission data access is fully VO)
- Built on International Virtual Observatory Alliance (IVOA) standards
- Recognised in the ESFRI roadmap (2018), supported in Europe by Euro-VO (VO Partners + EC projects since ~2001), recognised in ASTRONET roadmaps (2008, 2014, 2021)

ESO

SKAO
NWO-I-ASTRON

JIVE

CTAO

KIS
ORB

EGO

CNRS-OBAS
CNRS-CPPM

INAF

UEDIN

UHEI

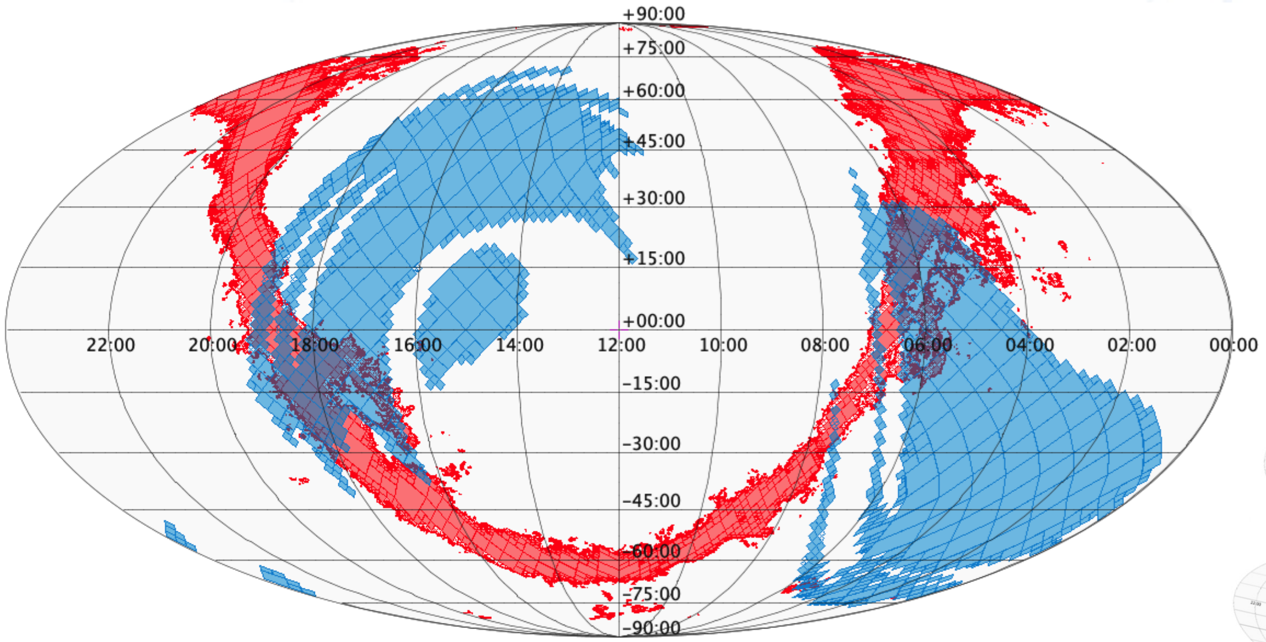
INTA

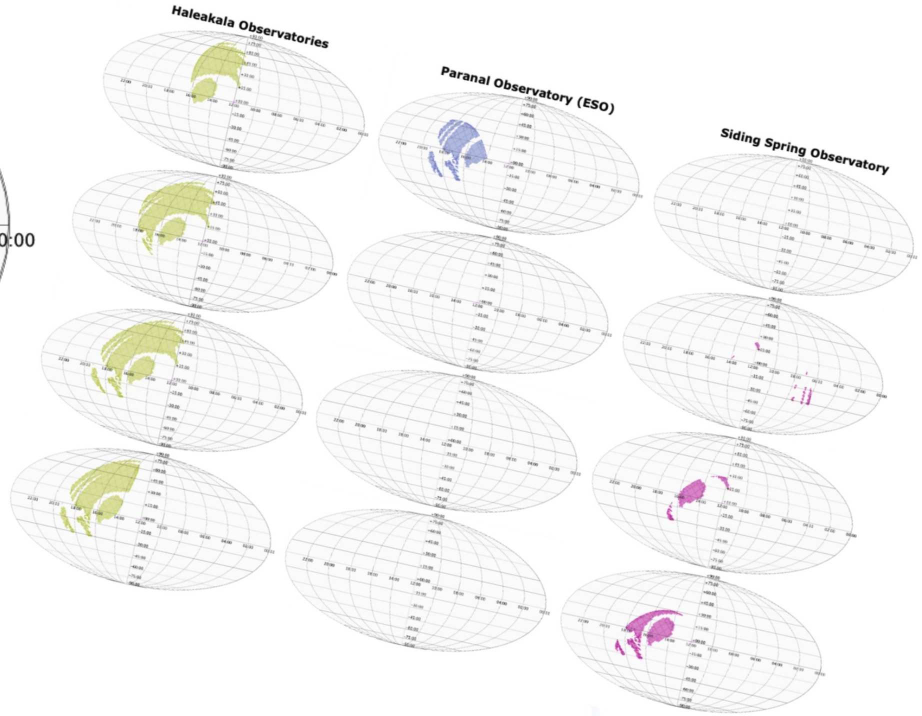
HITS (WP 3)

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

ESCAPE CEVO

- CEVO Examples - Planning of GW follow-up observations

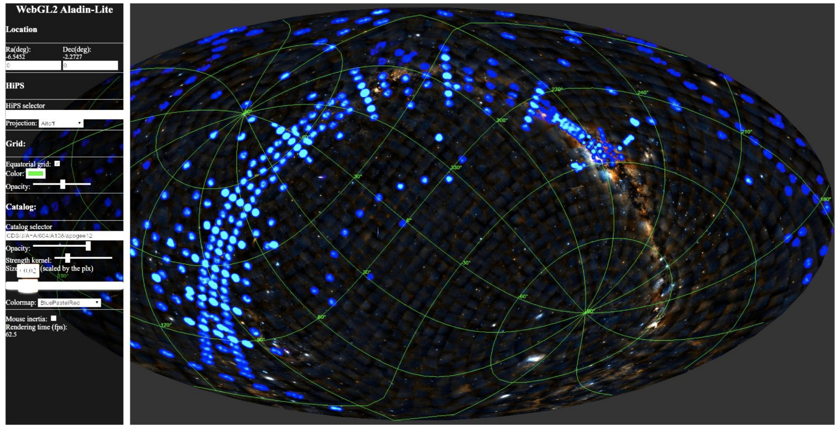




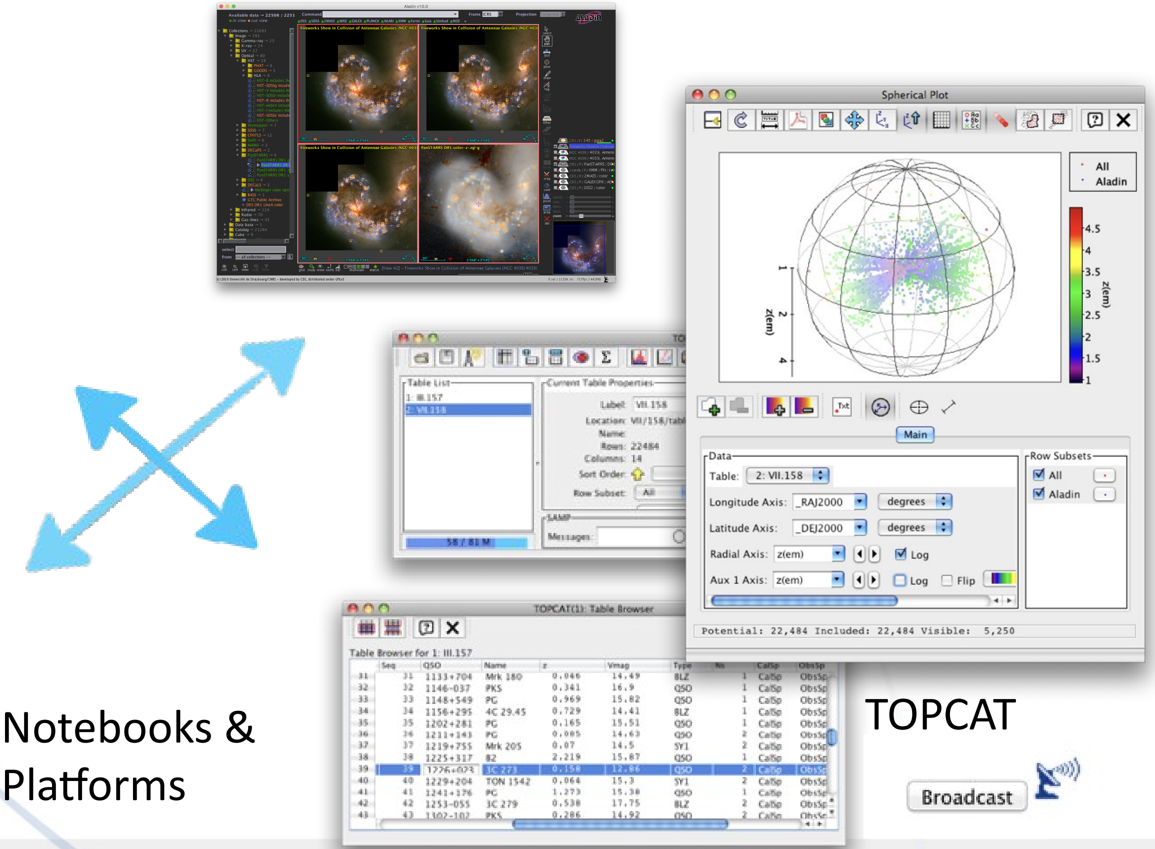
Funded by the European Union's
Horizon 2020 - Grant N° 824064

ESCAPE CEVO

- CEVO Examples – Aladin evolution, integration with analysis tools



New improved WebGL Aladin Lite



Notebooks & Platforms

TOPCAT

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

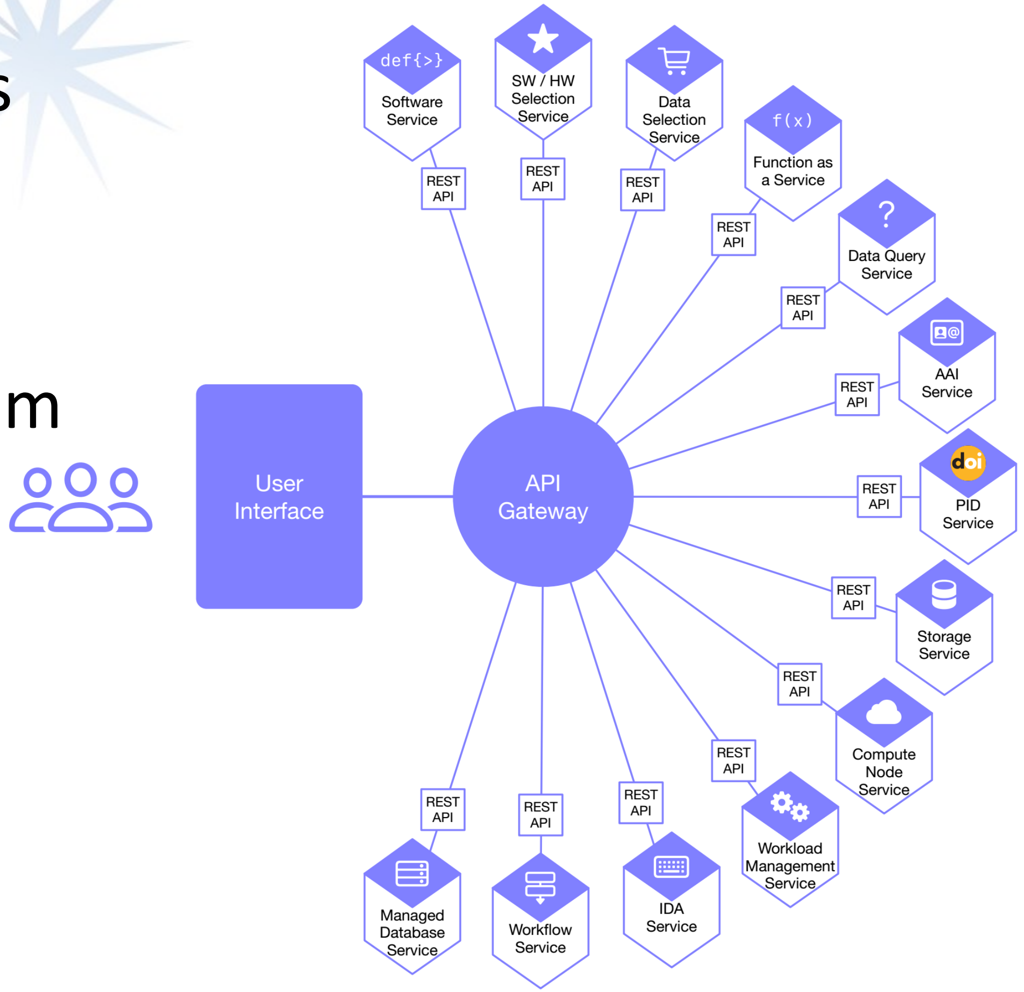
ESCAPE ESAP

ESFRI Science Analysis Platform

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

ESCAPE ESAP

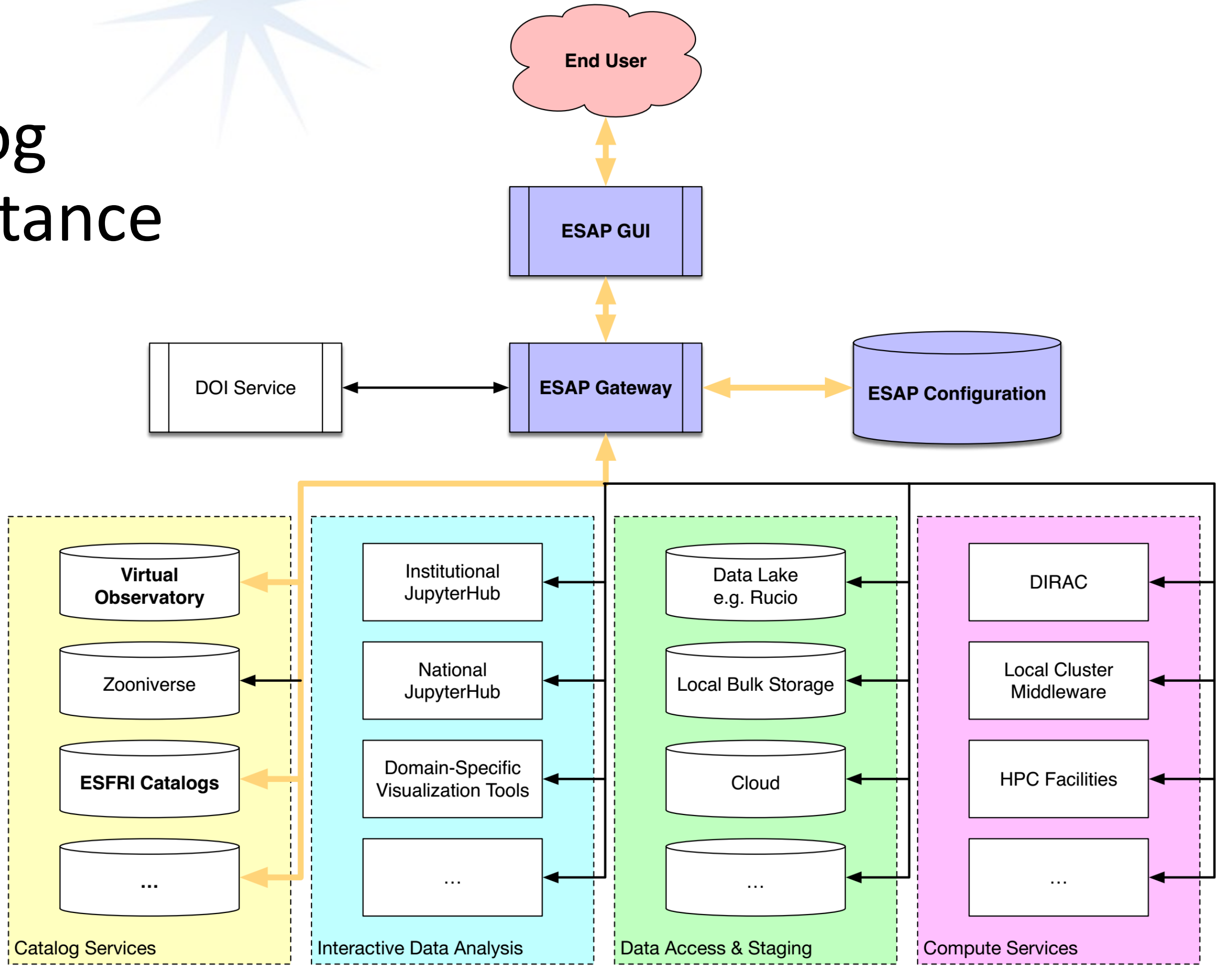
- ESAP: a focal point to integrate services from other providers
- Web UI and API gateway
- “Centralized ESAP”: access to a spectrum of ESCAPE services
- “ESFRI ESAP”: a way for individual infrastructures, projects, etc. to integrate diverse capabilities



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

ESCAPE ESAP

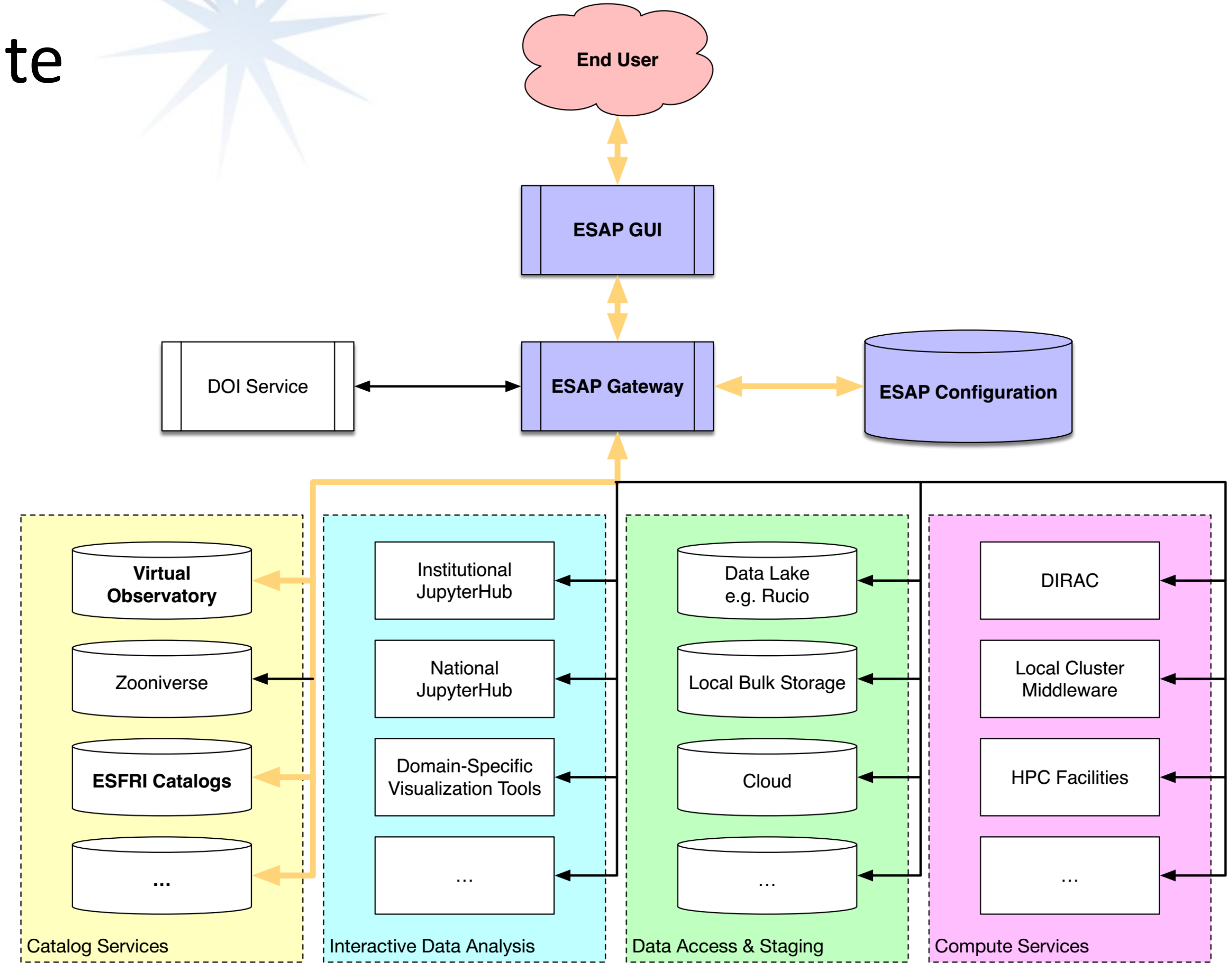
- ESAP Workflow example: query
  - User identifies relevant catalog services configured in this instance of ESAP
  - User submits search terms to multiple catalogs using consistent ESAP UI
  - Search results returned to user and displayed in unified form



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

ESCAPE ESAP

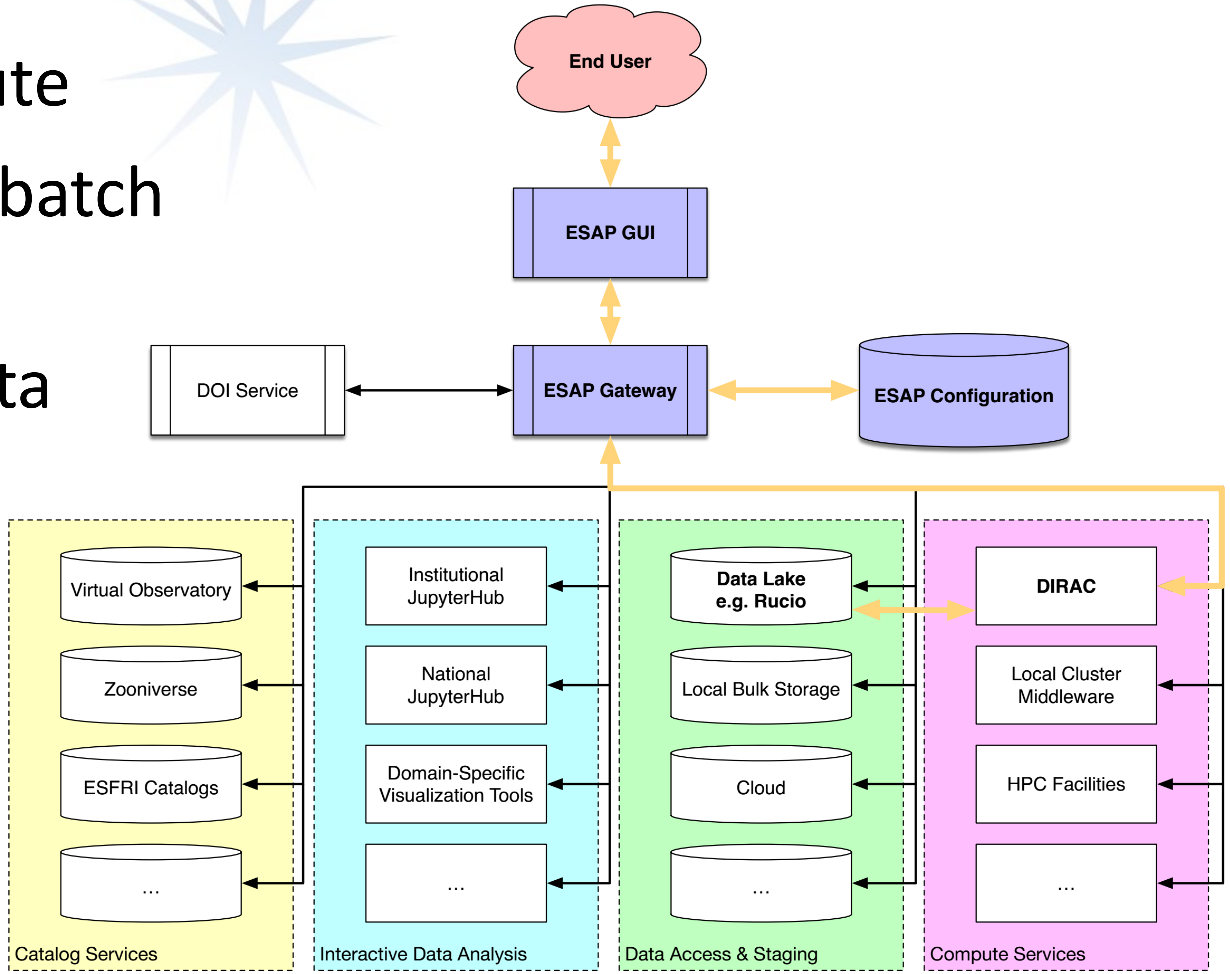
- ESAP Workflow example: compute
  - User identifies convenient Jupyter system configured in this instance of ESAP.
  - User sends retrieved catalog data to notebook
  - Interactive analysis session serves to prepare bulk processing



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

ESCAPE ESAP

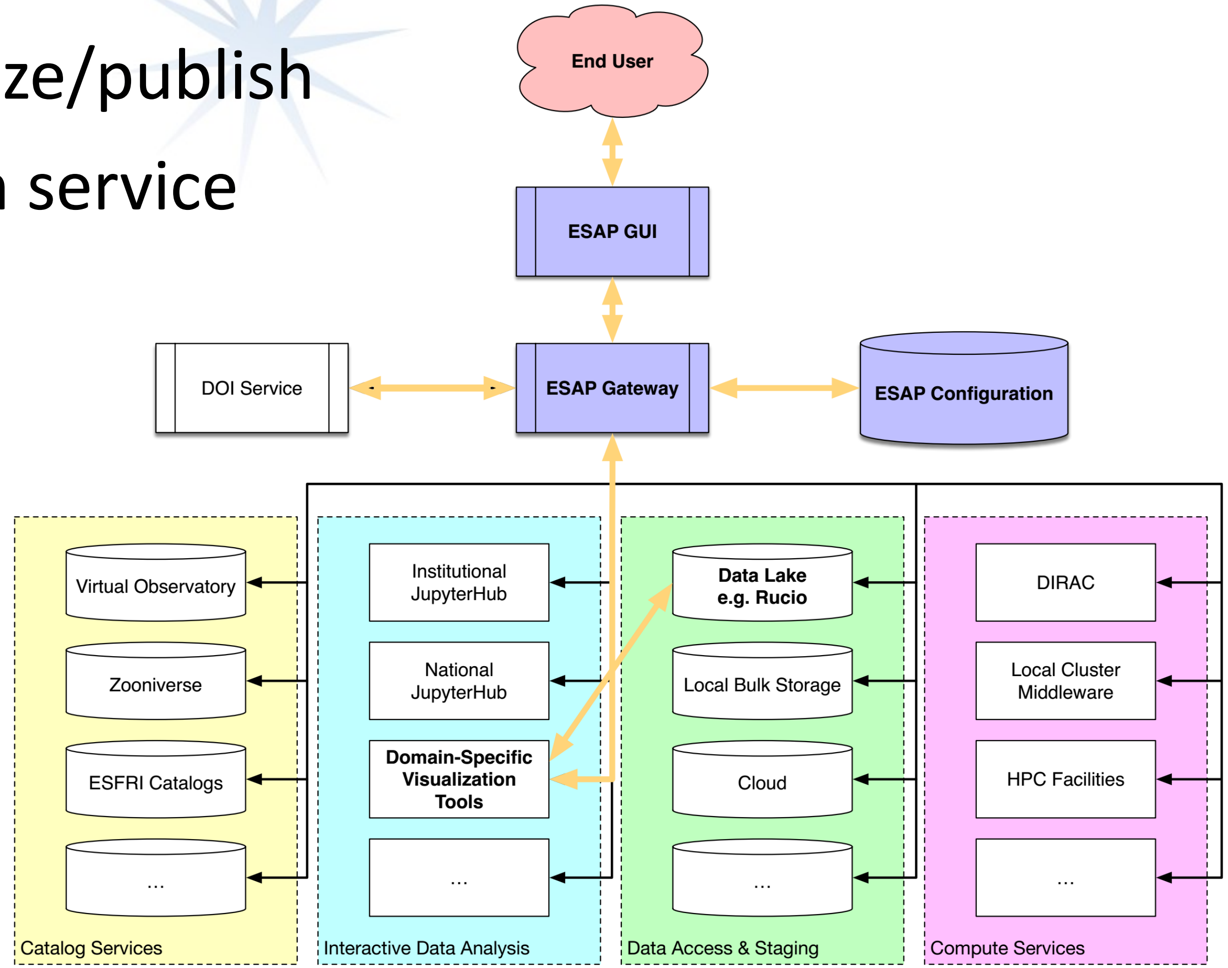
- ESAP Workflow example: compute
  - User sends instructions for a batch job to a DIRAC cluster
  - Compute cluster retrieves data from the Data Lake
  - Batch processing
  - Results to Data Lake and user notified of completion



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

ESCAPE ESAP

- ESAP Workflow example: visualize/publish
  - User identifies a visualization service
  - User initiates a visualization session, passing location of compute results
  - A DOI can be created and the results can be shared to a wider community



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

Test Science Projects

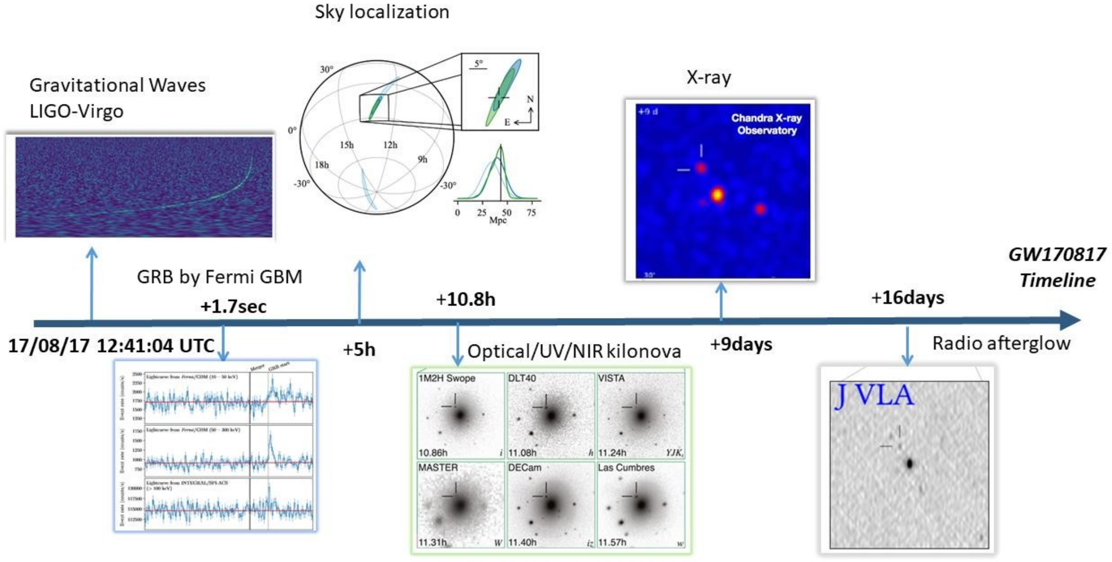
ESCAPE offspring: Test Science Projects

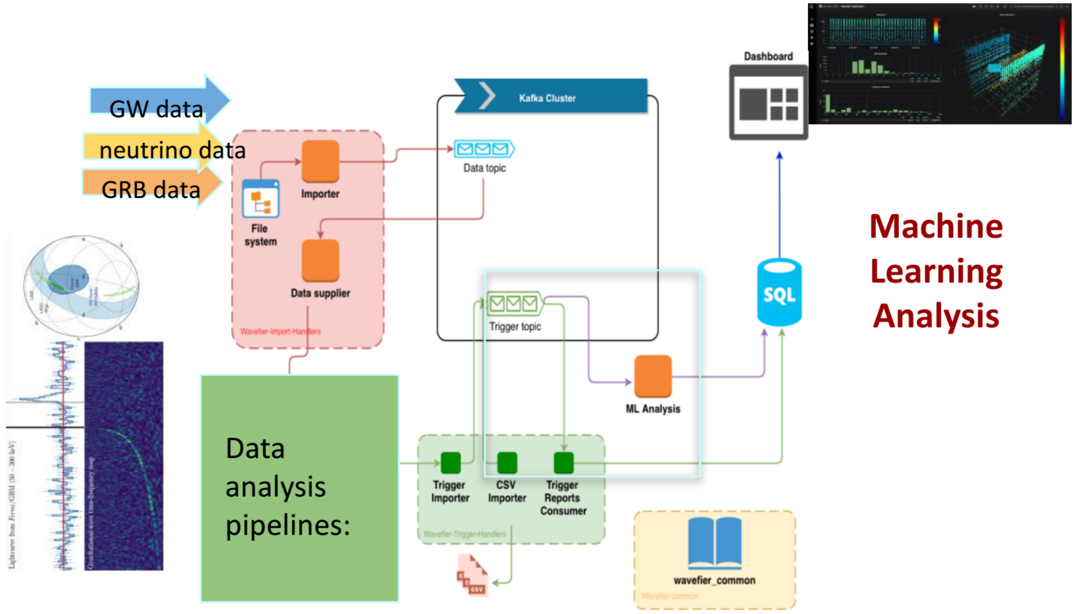
Funded by the European Union's Horizon 2020 - Grant N° 824064

Test Science Projects

- ESCAPE Test Science Projects: use the EOSC and ESCAPE-developed tools to set up multimessenger analysis platforms

Wavefier: a prototype for realtime analysis





Machine Learning Analysis

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

Test Science Projects

- ESCAPE Test Science Projects: use the EOSC and ESCAPE-developed tools to set up multimessenger analysis platforms
- Example: Extreme Universe TSP

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	Partner PM involved
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*	*42 PM Astron/UvA*
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*	*12 PM UvA, 6 PM FAU. 4 PM CNRS , 24 PM SNS*
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*	*10 PM INFN,12 PM UvA, 12 PM SNS, 2 PM FAU …*

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

Test Science Projects

- Example: Extreme Universe TSP/High Energy Astrophysics

Stage 1

Wavefier running at CNAF

Wavefier linked to ESCAPE data lake

Stage 2

CTA simulated data

Wavefier with 2 streams of data:
Release of GW-Wavefier on Zenodo + Wavier documentation.

Stage 3

KM3NeT data

Wavefier with 3 streams of data

Stage 4

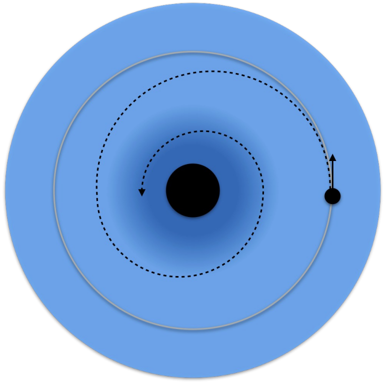
Multiprobe-Wavefier

Machine learning pipeline for 3 streams of data.

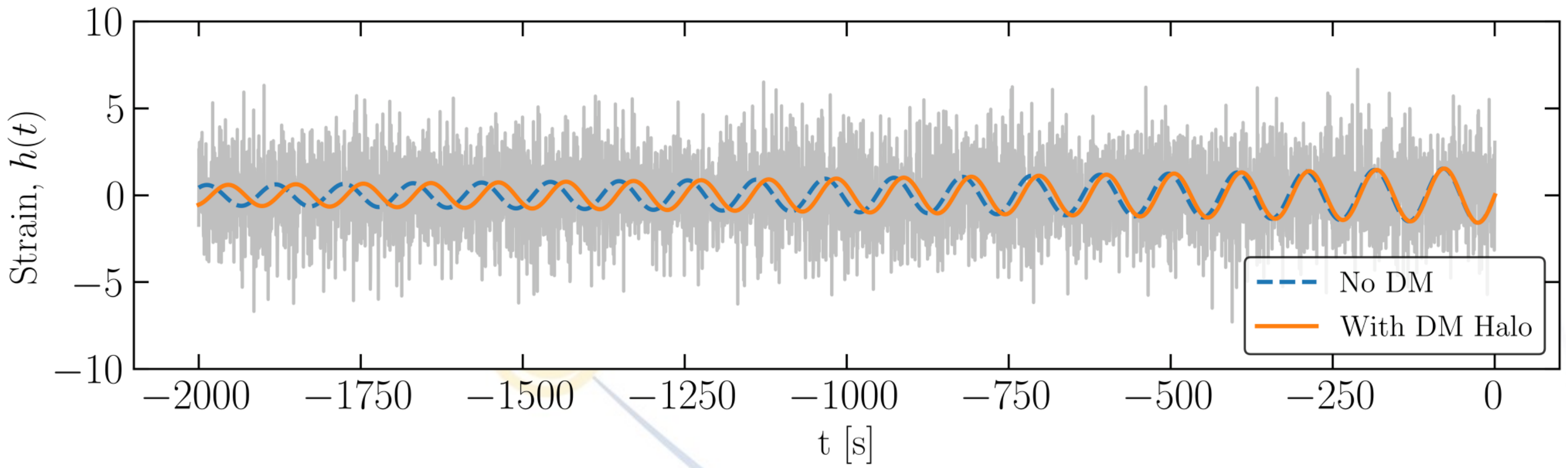
Funded by the European Union's Horizon 2020 - Grant N° 824064

Test Science Projects

Example: Extreme Universe TSP/Detecting DM with the Einstein Telescope



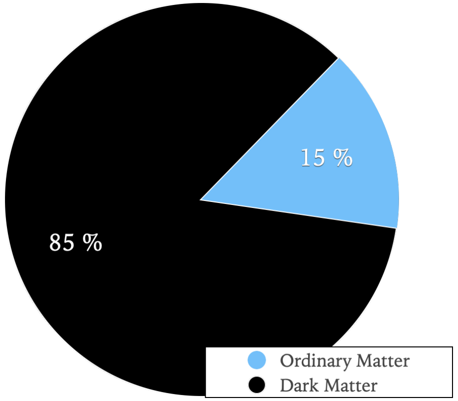
- Presence of DM ‘spikes’ around BHs can alter inspiral dynamics
- GW waveform gradually goes out of phase with the corresponding vacuum-only waveform
- Possibility to detect and constrain dense DM ‘spikes’ with just a few cycles of GW ‘dephasing’ → but these subtle differences can be hard to detect!

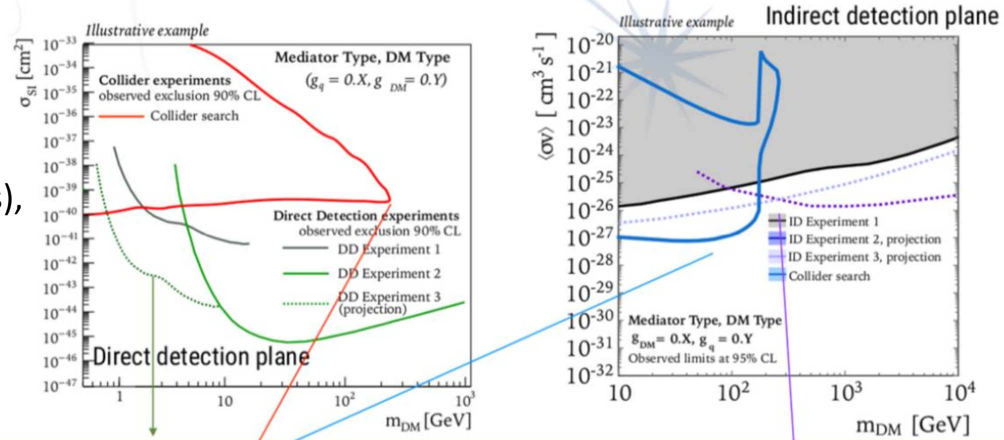


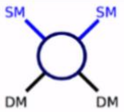
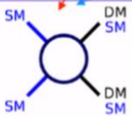
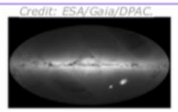
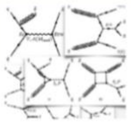
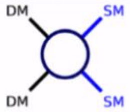
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Test Science Projects

- # Example: Dark Matter
  - Implement end-to-end dark matter analysis workflows on the EOSC, taking into account diversity in datasets (from kilobytes to petabytes), workflows (Jupyter notebook vs Big Data analysis), experimental communities (from individual theorists to large international labs)



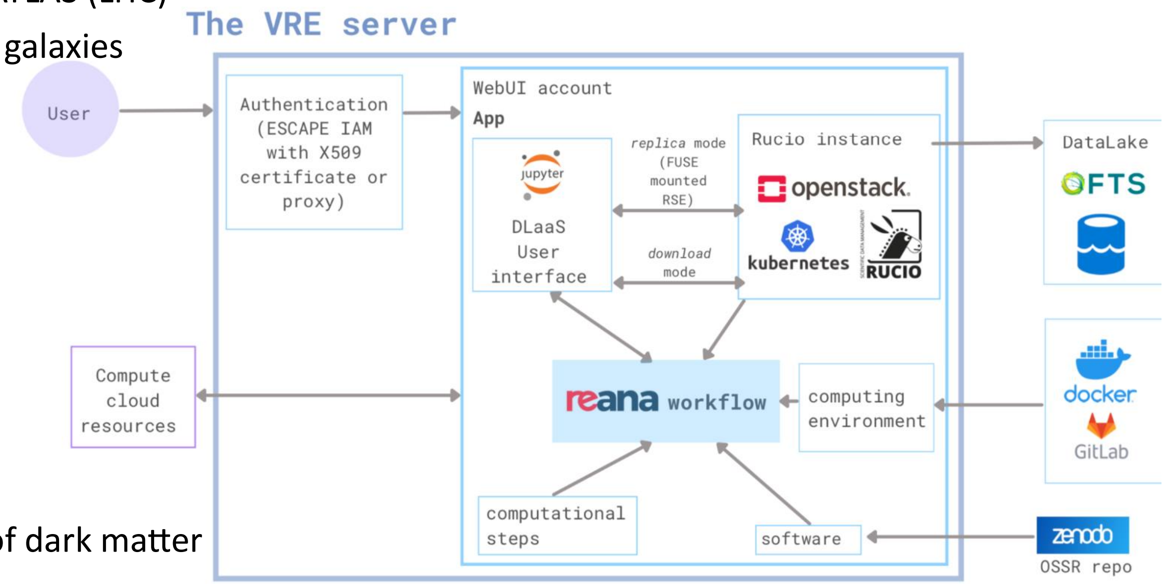


METHODS	 Direct Detection	 Colliders	 Astrophysics	 Theory	 Indirect Detection
RELEVANCE	DM that interacts inside the detector (WIMPs, axions)	produce DM and probe the dark interaction	necessary for all	necessary for all	detect annihilating/ decaying DM through its decays (i.e. neutrino searches, gamma rays)
EXPERIMENTS INVOLVED	Darkside	ATLAS			KM3Net, CTA

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

Test Science Projects

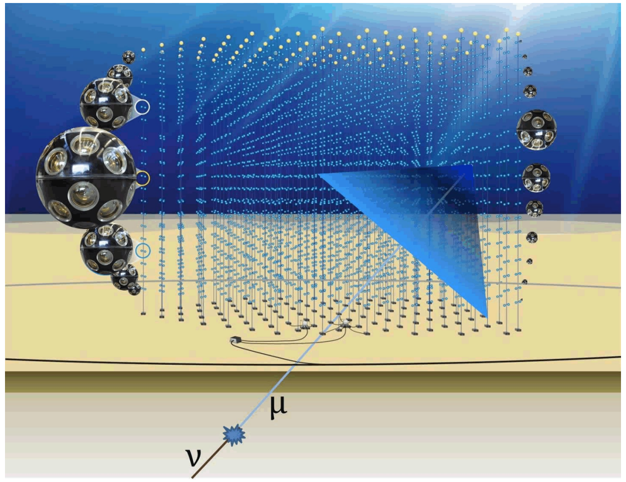
- ## Example: Dark Matter
- Implement end-to-end dark matter analysis workflows on the EOSC, taking into account diversity in datasets (from kilobytes to petabytes), workflows (Jupyter notebook vs Big Data analysis), experimental communities (from individual theorists to large international labs)
- Onboarded onto the VRE (Virtual Research Environment)
  - Reinterpretation of search for particles mediating dark matter at ATLAS (LHC)
  - Analysis of Fermi-LAT data for dark matter constraints from dwarf galaxies
  - Analysis of KM3NeT experiment data for DM indirect detection
- To be onboarded onto the VRE
  - Two new full ATLAS/LHC searches for dark matter mediators
  - Real-time analysis
  - Leptons + missing transverse momentum
  - A reinterpretation of CMS search for dark matter production at the LHC
  - Compression algorithms using Machine Learning
  - Simulation pipeline for DarkSide experiment for direct detection of dark matter



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

Test Science Projects

- Example: IRF from KM3NeT for Dark Matter and Extreme Universe TSPs



- Instrument Response Function of neutrino telescope provides a quantitative estimation of the event rate
- It allows avoiding extensive MC simulations each time for a new configuration of neutrino source
- Different configurations of neutrino sources:
  - Diffuse source
  - Point source with power law E^{-a}
  - Extended source
- Using gammapy interface will give an easily combination with other gamma experiments like CTA

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

What’s next

ESCAPE Collaboration Agreement

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

ESCAPE Collaboration Agreement

- ESCAPE will become a sustained "Community Platform": (ESF)RIs core partners as Parties in a new Collaboration Agreement to operate Open Science and cooperate to address new topics in ERA

[…]

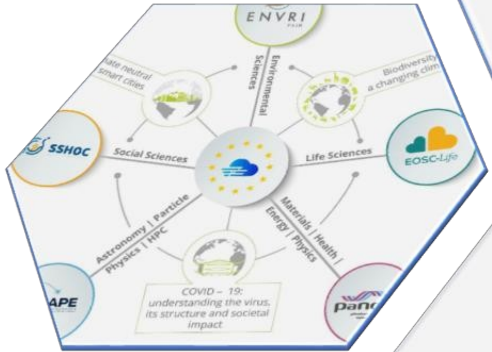
Funded by the European Union's Horizon 2020 - Grant N° 824064

ESCAPE Collaboration Agreement

For EOSC destination and	European Research Area (ERA) operational
Data Lake	Low Latency data analysis
Software catalogue, methods and training	Technological dev. and innovation (e.g. quantum)
European Virtual Software Institute for Research	Curiosity driven RIs' synergies
Virtual Research Environments	Inter-cluster projects for society
Open Science Research Projects	Citizen Science
Data interoperability	Researchers' rewarding
	Sector Data Spaces [...]

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

ESCAPE Collaboration Agreement

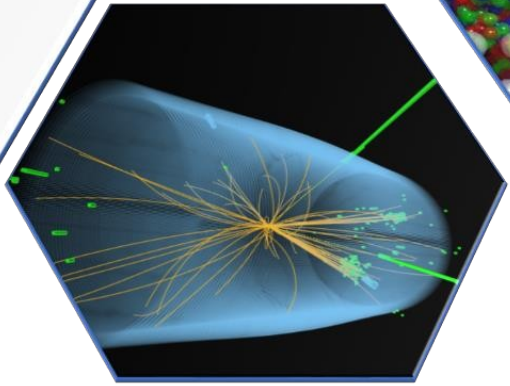


Community-based scientific software development and preservation

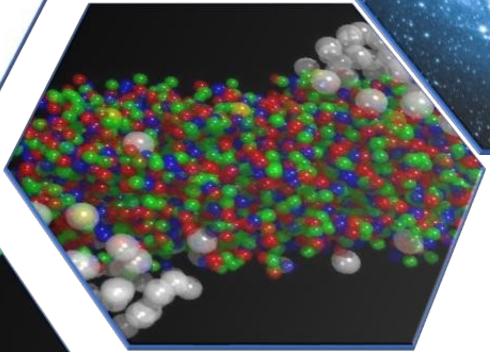
Advancing frontier knowledge of our RIs

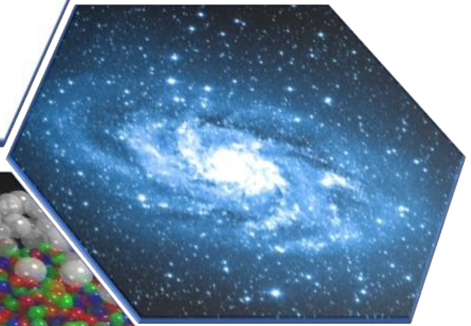
VRE for assessing scientific knowledge production

Build on the Science Clusters approach to ensure the uptake of EOSC



Future shared development of RI technologies and services





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