ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

ASTRONOMY & PARTICLE PHYSICS CLUSTER

Mark ALLEN

ASOV Annual Meeting, Paris 10-11 March 2020

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.

ESCAPE

- An overview of the status of ESCAPE
  - from the ESCAPE Progress Meeting, 26-27 Feb 2020
    - ESCAPE as one of the “cluster projects”
    - EOSC – from EOSC secretariat / Architecture WG
    - Progress in the various WPs
  - WP4 – connecting to EOSC via the VO
    - Highlights from first year
  - various slides are from presentations made by participants at the ESCAPE Progress meeting

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

ESCAPE in a nuthshell

- **31** partners (including 2 SMEs)
- **7** ESFRI projects & landmarks: CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA
- **2** pan-European International Organizations: CERN, ESO (with their world-class established infrastructures, experiments and observatories).
- **3** supporting European research consortia: APPEC, ASTRONET and NuPECC.
- **1** involved initiative/infrastructure: EURO-VO
- **2** European research infrastructures: EGO and JIV-ERIC
- Started: **1/2/2019**
- Duration: **42** months (end date 31/7/2022)
- Grant number: **824064**
- Coordinator: **CNRS-LAPP**

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

The role of the European Open Science Cloud (EOSC) is to ensure that **European scientists reap the full benefits of data-driven science**, by offering:

1.9 million *European researchers and* ***70 million*** *professionals in science and technology a* ***trusted open distributed environment*** *providing seamless access to data and interoperable services addressing the whole research data life cycle.*

The development of the EOSC realises EU policy objectives including Open Science, FAIR data implementation and the Digital Single Market

European Commission

What is EOSC?


Trusted digital platform
FAIR (Findable, Accessible, Interoperable Reusable) standards
Innovative services
Universal gateway
Federation
EOSC
User-oriented
Across borders
Rules for participation
Seamless access
Cross-disciplinary research

Why EOSC?

European Commission

EOSC today

EOSC in numbers

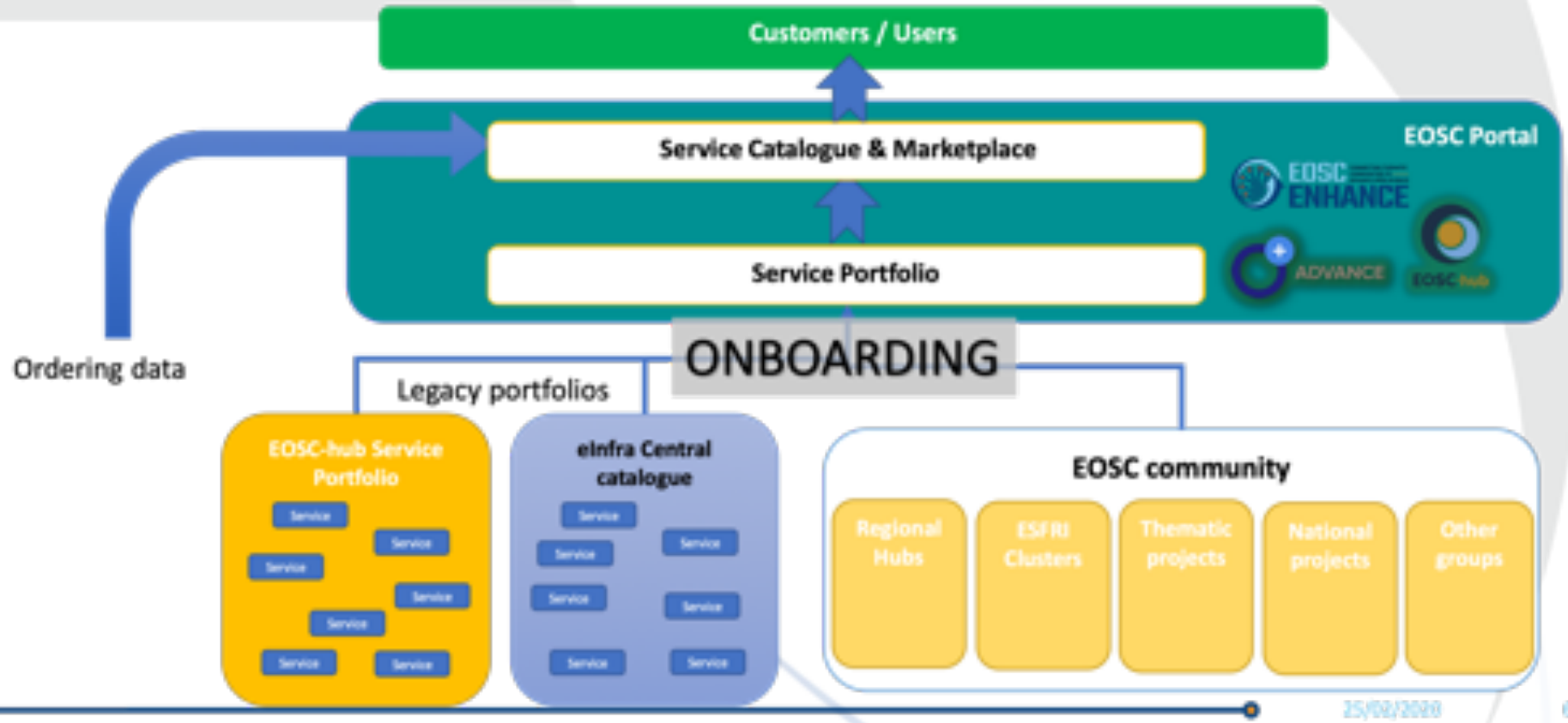
1.9 million researchers and **70 million** professionals will benefit from EOSC

€ 250 million in 2018-2020 (Horizon 2020)

>30 core EU-funded projects implementing EOSC

The EOSC implementation roadmap, published in 2018 by the Commission, addresses the implementation of EOSC under six action lines:

EOSC Portal 2020



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

EC cluster-projects initiative

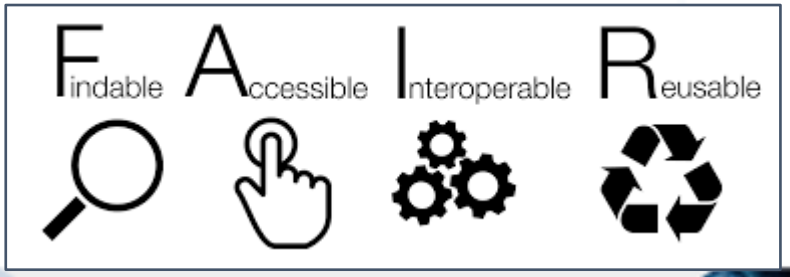
H2020-INFRAEOSC-04-2018 call

Clusters to ensure the connection of the ESFRI RIs with EOSC (and the construction of EOSC)

Expected impact:

- *Improve access to data and tools leading to new insights and innovation*
- *Facilitate access of researchers to data and resources for data driven science.*
- *Create a cross-border open innovation environment.*
- *Rise the efficiency and productivity of researchers through open data services and infrastructures for discovering, accessing, and reusing data.*
- *Foster the establishment of global standards.*
- *Develop synergies and complementarity between involved research infrastructures.*
- *Adopt common approaches to the data management for economies of scale.*

Making data FAIR …



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

Astronomy and Particle Physics

- The astronomy-related ESFRI projects and the accelerator-based particle physics ESFRI facilities will open together new paths towards the understanding of the Universe through a multi-probe approach.

- Enhance the coordination leveraging two major complementary excellences in data stewardship:

  i) the astronomy Virtual Observatory infrastructure;

  ii) long-standing expertise of the particle physics community in large-scale distributed computing and big-data management.

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

ESCAPE ESFRI facilities aligned expectations

The “Data-FAIRness” challenge for our research infrastructures implies:

- Large volumes of data generators (up to multi-Exabyte scale level) -> Not only early adapters of the latest ICT and data-management developments but also constantly pushing the envelope of the current state-of-the-art.

- “Observatory” and “Facility” type of operation requires global open access and long-term sustainability of *FAIR* research data and services.

- Training and extension of FAIRness standards and tools for data access and data preservation.

- Operating a common/transversal virtual research environment, sharing competence, adopting the Open-Science principles and boarding EOSC.

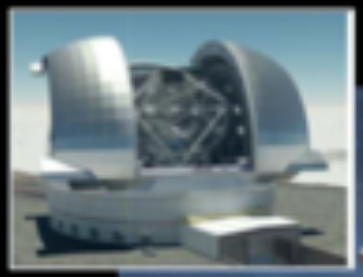
Funded by the European Union’s Horizon 2020 - Grant N° 824064

Radio

SKA

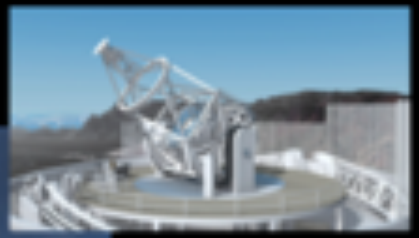
JIVE-VLBI

Visible light



ELT

ESO



EST

Gamma rays

CTA

Accelerator-based
Particle Physics

HL-LHC

CERN

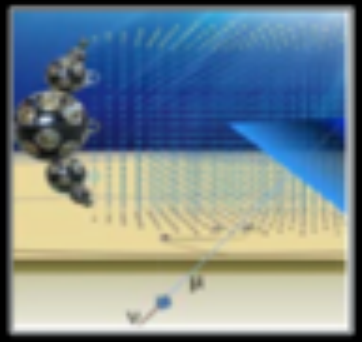
Accelerator-based
Nuclear Physics

FAIR

Gravitational
Waves

EGO-VIRGO

Cosmic-rays
Neutrinos



KM3NeT

ESCAPE goals towards a VRE

1. Implementing Science Analysis Platforms for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools, bring their own custom workflows.

2. Contributing to the EOSC global resources federation through a Data-Lake concept implementation to manage extremely large data volumes at the multi-Exabyte level.

3. Supporting "scientific software" as a major component of ESFRI data to be preserved and accessed in EOSC through dedicated catalogues.

4. Implementing a community foundation approach for continuous software shared development and training new generation researchers.

5. Extending the Virtual Observatory standards and methods according to *FAIR* principles to a larger scientific context; demonstrating EOSC capacity to include existing frameworks.

6. Further involving society in knowledge discovery.

ESCAPE SAP Science Analysis Platform

ESCAPE DIOS Data Infrastructure for Open Science

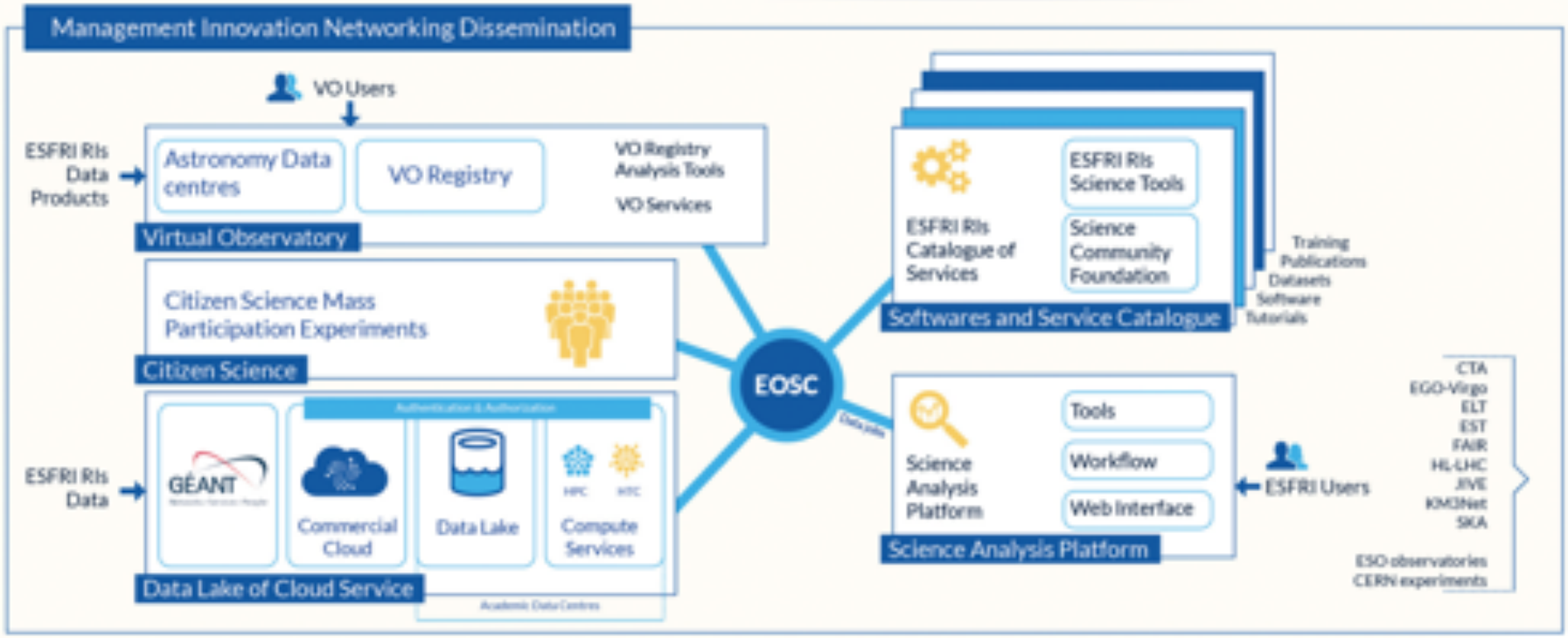
ESCAPE OSSR Open-source Scientific Software and Service Repository

ESCAPE VO Virtual Observatory

ESCAPE CS Citizen Science

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

ESCAPE goals: building a domain-based EOSC cell



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

External Advisory Board – Particle Physics and Astrophysics Networks + ESA

APPEC
Teresa Montaruli - Chair

ASTRONET
Colin Vincent - Chair

NuPPEC
Marek Lewitowicz - Chair

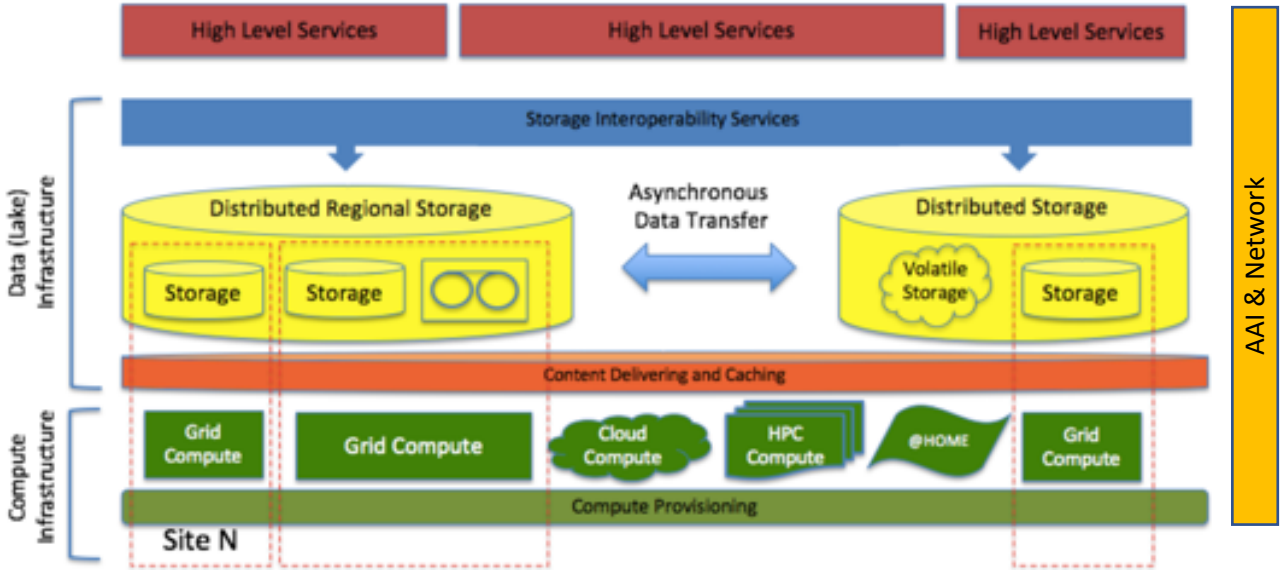
ECFA
Jorgen D'Hondt - Chair

ESA
Christophe Arviset

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

WP2 – Data Infrastructure for Open Science

Data Lake building blocks



Define, integrate and commission an ecosystem of tools and services to build a data lake

Leaves to the science projects the flexibility to choose the services and layout most suitable to their needs. Provides a reference implementation

Contributes to deliver Open Access and FAIR data services: relies on trustable data repositories; enables data management policies; hides the complexities of the underlying infrastructure providing a transparent data access layer

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

WP2 – Progress and next steps

Year 1: we have a Data Lake pilot, well ahead of the milestone deadline

- Architecture driven by the ESCAPE sciences
- Technology in synergy and leveraging existing projects and solutions

Year 2: demonstrate its usability for the ESCAPE sciences

- Organize and manage experiment data
- Process such data integrating compute platforms

Year 2 and beyond: scale up the deployment. Introduce advanced features.

WP3 - E-OSSR Aims and Objectives

- Aim: expose the tools of the ESCAPE (ESF)RI projects in a repository under the EOSC catalogue of services
- Objectives:
  - continuous development, deployment, exposure and preservation of software/tools/services
  - interoperability, software re-use and cross-fertilisation
  - open innovation environment for open standards (workflows), common regulation and shared (novel) software for multi-messenger&multi-probe data
- All objectives follow:
  - a community-based approach
  - the FAIR principles for open software/services and data
- E-OSSR strives to:
  - Establish a foundation to (co-)develop EOSC-ready software and services;
  - expose them to users via the EOSC catalogue of services;

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

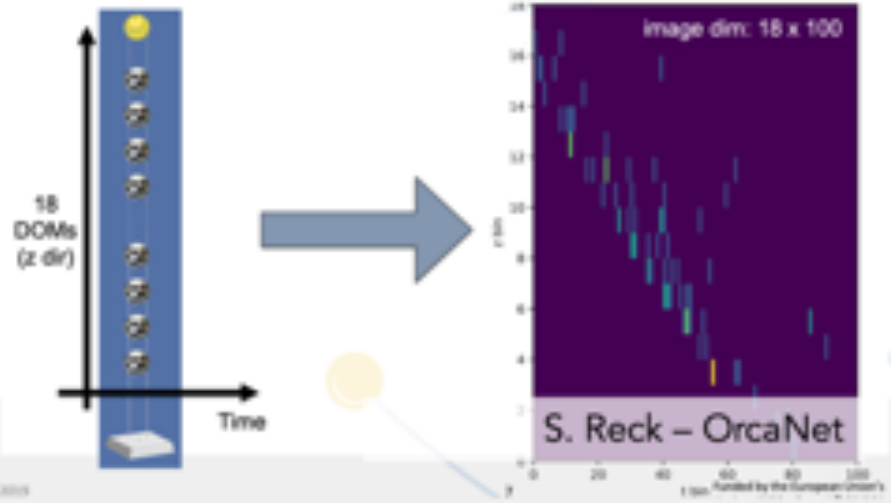
Results from the Focus Groups: FG1: Collecting Software Requirements

- Started development, benchmarking and deployment of software within and across partner institutes;
  - template for software gathering set up;
  - first round of software to become part of the repository collected;
  - partners started to prepare the software for repository integration;
- Preliminary outcome/Deliverable:
  - Software and service list and integration plan (deliverable D3.2)
  - software will be linked to the repository with source code & a containerised solution with test data and documentation for local tests
- Next steps:
  - How to make software discoverable by repository (meta data)?
  - gathering of common practices and know-how towards the definition of best practices to be shared with the community;
  - pick use cases for repository integration

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

Results from the Focus Groups: FG3: Innovative workflows

- Machine learning approaches to simulation and experiment data adapted and benchmarked;
- Definition of data formats and different deep-learning approaches pursued;
- Exchange of experience, gathering and harmonisation of approaches for innovative workflows between different partners;
- Next steps: establishment of test science cases for multi-messenger analysis workflows connecting several ESFRIs
  ⇒ cross-fertilization across WPs, define requirements via use cases



Imaging gamma rays

Th. Vuillaume -

Work Package 4 – WP4 “CEVO”

Connecting **E**SFRI projects to EOSC through the **VO** framework

- **EOSC** – **E**uropean **O**pen **S**cience **C**loud
- **VO** – **V**irtual **O**bservatory

Virtual Observatory standards and methods for FAIR principles to a larger scientific context; demonstrating EOSC capacity to include existing frameworks.

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

WP4 Tasks

Task 4.1 Integration of astronomy VO data and services into the EOSC

Lead: Marco Molinaro (INAF)

Task 4.2 Implementation of FAIR principles for ESFRI data through the Virtual Observatory

Lead: Françoise Genova (CNRS-ObAS)

Task 4.3 Adding value to trusted content in astronomy archives

Co-leads: Mark Allen (CNRS-ObAS) & Martino Romaniello (ESO)

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

Progress in year 1 – *all tasks active*

Task 4.1 Integration of astronomy VO data and services into the EOSC

- **Interaction with EOSC bodies, VO registry in B2FIND, tests of service publishing**

Task 4.2 Implementation of FAIR principles for ESFRI data through the Virtual Observatory

- **Milestones – representation of ESCAPE priorities at IVOA level**
- **ESFRI/RI partners requirements defined, results on tools and VO publishing**

Task 4.3 Adding value to trusted content in astronomy archives

- **First results of machine learning applied to archive data sets**

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

WP4 specific events in year 1

- Transition event – ASTERICS Tech Forum (Feb 2019)
- Radio Astronomy and VO meeting / EST and VO meeting (Feb 2019)
- VO partner visits to ESFRI/RIs: **EST**: ROB & KIS, ASTRON, KM3NeT
- EOSC-Hub Week (April 2019)
- KM3NeT and VO meeting – September 2019
- Task 4.3 Meetings (September, December 2019)
- Provenance – CTA, KM3NeT (Nov 2019)
- SCIOPS – presentation of ESCAPE (Nov 2019)
- EOSC Symposium (Nov 2019)
- WP4 Technology Forum 1 (Feb 2020)

IVOA Paris
May, 2019

IVOA Groningen
October, 2019

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

What is the Virtual Observatory?

- **Operational framework** for interoperable access to astronomical data and services across all areas of astronomy

- Provides unique scientific capabilities, opening up new ways of using rich data in astronomy archives and services

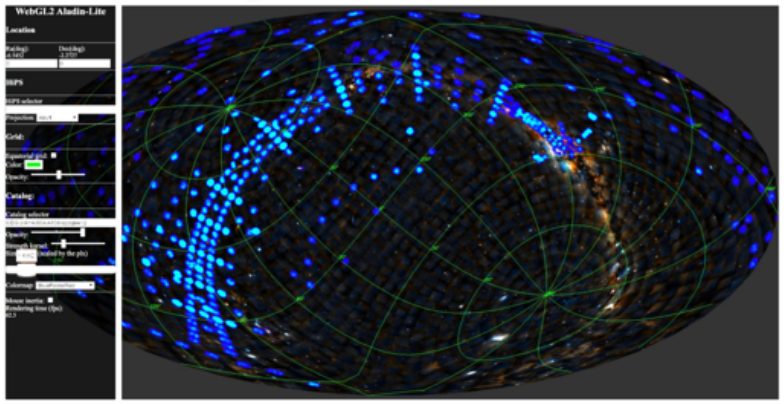
- **A pioneer of FAIR data sharing** - **an existing global framework** – populated by major data providers (space and ground based) that is heavily used by the community *(e.g. Gaia data access is fully VO)*

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

One view of the VO from an application:

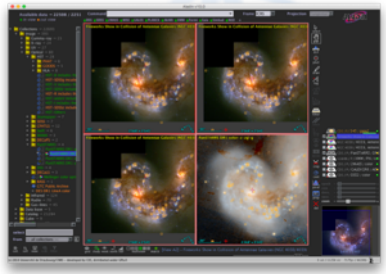
Funded by the European Union's Horizon 2020 - Grant N° 824064

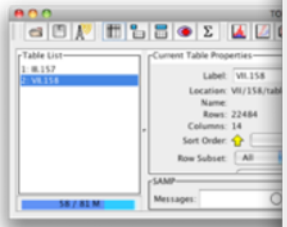
CEVO: VO interoperability in context of ESCAPE / EOSC

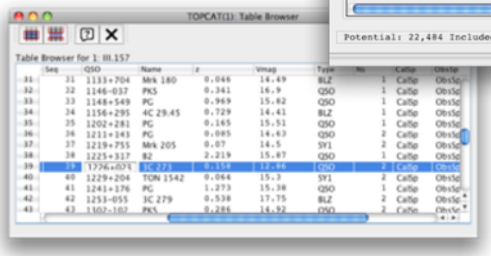


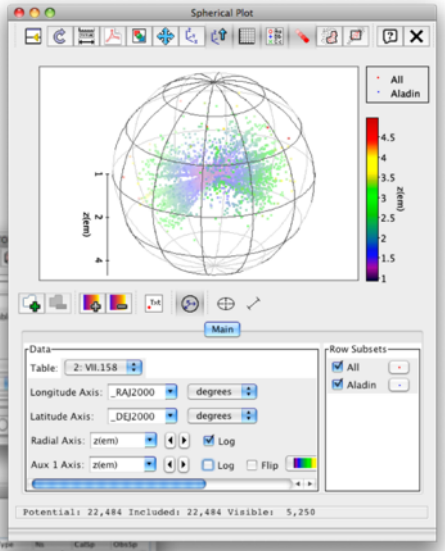
New improved WebGL Aladin Lite

Notebooks & Platforms









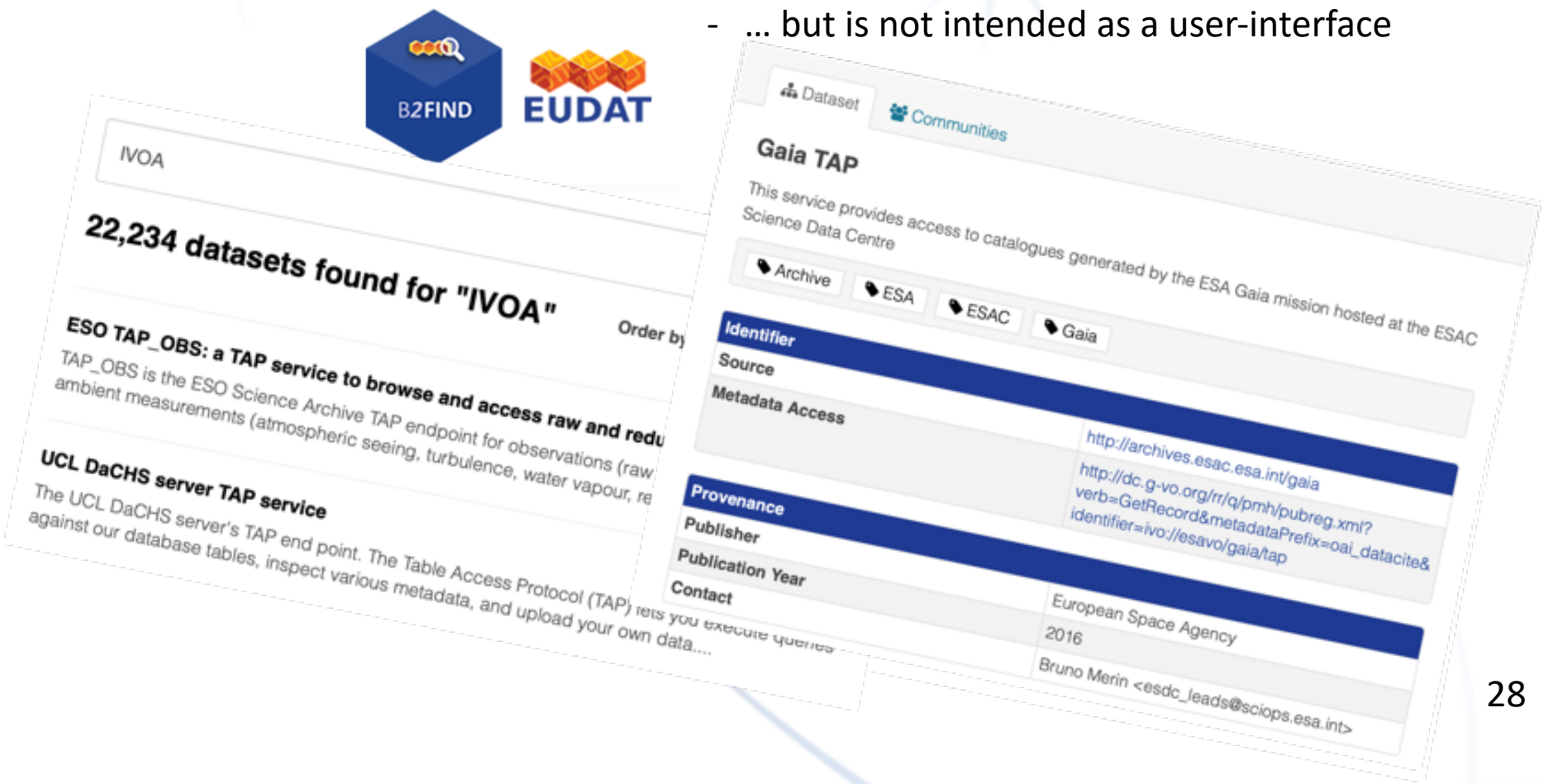
TOPCAT

Broadcast

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

Task 4.1 B2FIND - Demonstrates 1st level of metadata compatibility

- Links to the actual service
- … but is not intended as a user-interface



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

Task 4.2 – IVOA Milestones

*Common domain-specific standards for **FAIR** data*

- First Milestone – **Paris IVOA meeting** *(May 2019)*
  - Introduction of solar physics requirements at IVOA, EST partners
  - EUDAT participation – registry integration
  - **30** contributions from ESCAPE partners, ESCAPE highlighted at IVOA level
- Second Milestone – **Groningen IVOA meeting** *(November 2019)*
  - Focus on Radio/mm astronomy – SKAO, LOFAR, JIVE,ALMA + international radio projects -- leading to IVOA Interest Group on Radio Astronomy
  - Space-time indexing of astronomy data
  - **22** contributions from ESCAPE partners

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

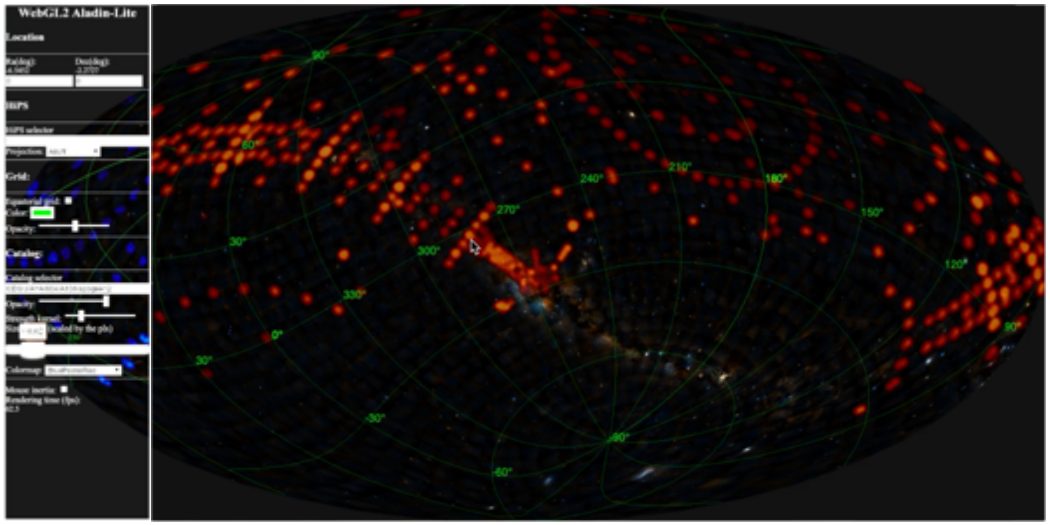
WP4 Technology Forum 1

- Major Work Package meeting (+ WP3, WP5 interaction)
- Invited experts and collaborators
  - **EUROPLANET, TOPCAT-**developer**, ObsParis – Radio & Solar expertise, ESA** *(Videocon)*
- Progress presentations + *Hack-a-thon* working meeting sessions

- **Mapping VO expertise – ESFRI needs**
- **Live demos and coding**
- **Tracking progress**

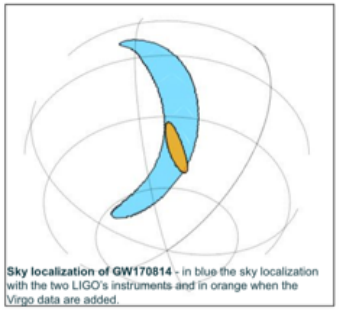
Funded by the European Union's Horizon 2020 - Grant N° 824064

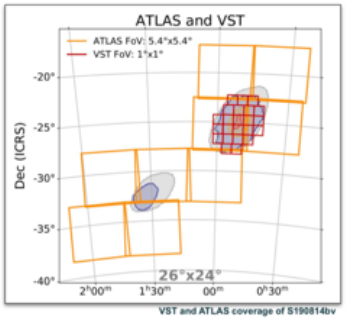
e.g. progress on data access and visualisation

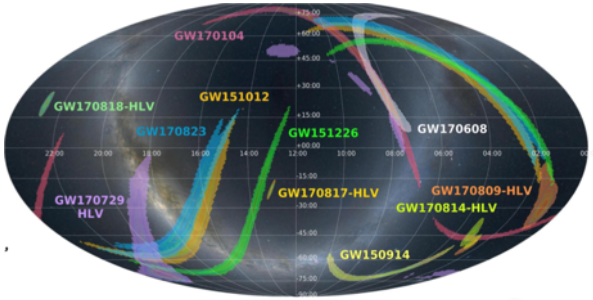


Prototype WebGL-enabled visualization component (Aladin Lite) for portals / platforms

IVOA MOC standard adapted for GW sky localization



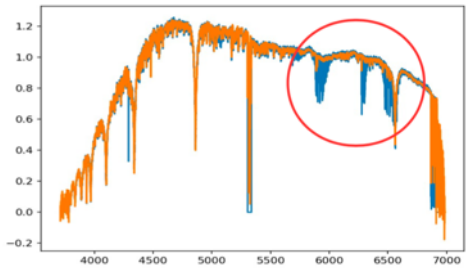




Funded by the European Union's Horizon 2020 - Grant N° 824064

Task 4.3 – Deep Learning and Data Archives

- Scope: provide archive users with novel ways to identify data
  - Beyond query parameters approach – *Let the data speak!*
  - Extension of recent move from instrument to data keywords (e.g. exposure time to signal-to-noise
- Deep Learning analysis of entire ESO HARPS archive (~270k spectra)
  - *Search for 'similar spectra'*
- Different approaches
  - **HITS**: fully-connected autoencoder 2 latent d
    - Agile architecture for speed and interactivity
  - **ESO**: combination of convolutional and fully-connected layers with between 4 to 8192 latent dimensions, full spectral resolution
    - Find minimal representation which preserves all the relevant information



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

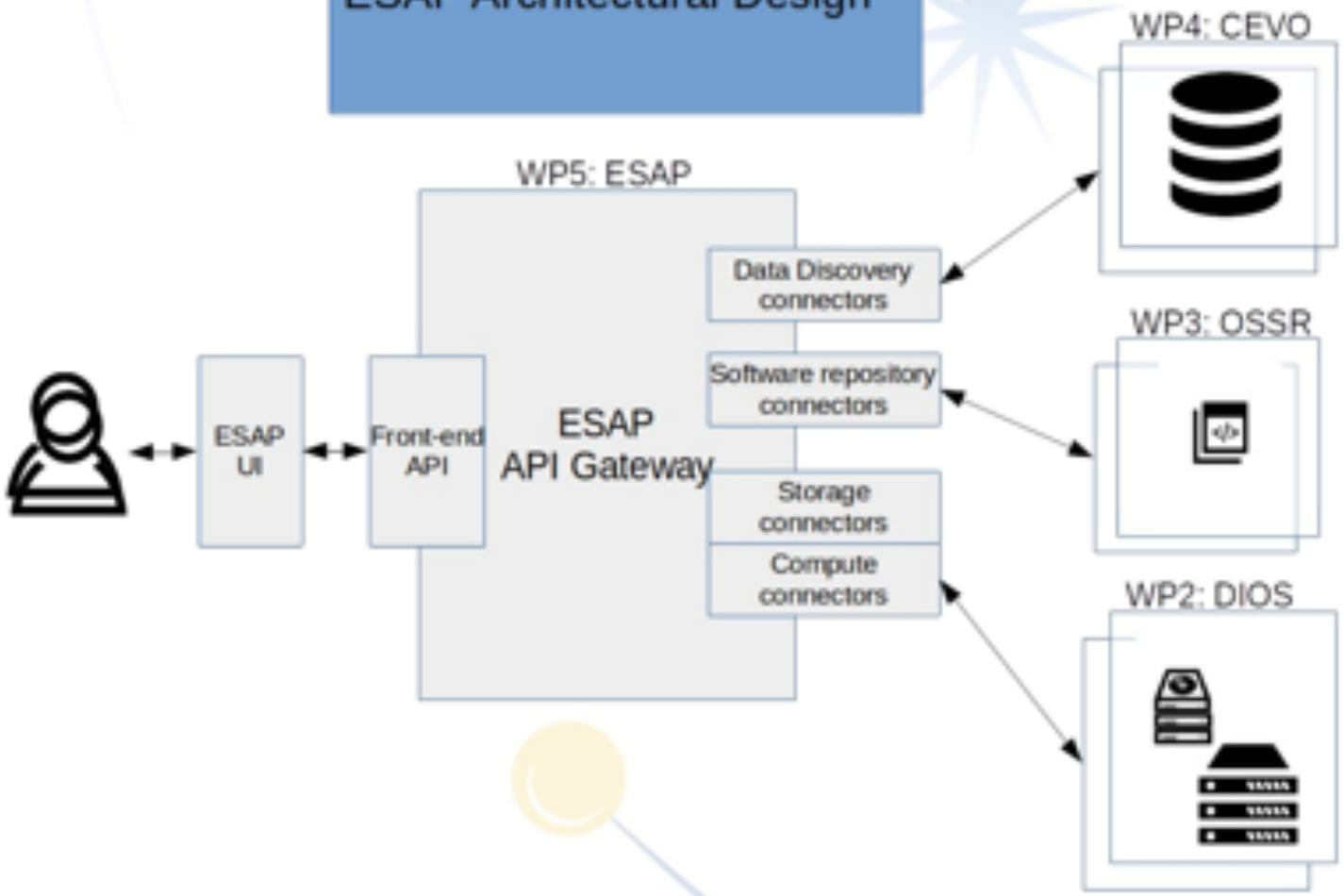
WP5 - ESCAPE Science Analysis Platform (ESAP)

- AAI
- Data selection shopping cart
- Data Staging and Access
- List of suggested Sofware/Workflows
- List of suggested Compute Resources
- Batch data processing
- Interactive Data analysis
- Data analysis with visualisation tools
- Data analysis with machine learning tools
- Research object catalogue
- Ingestion of advanced data products

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

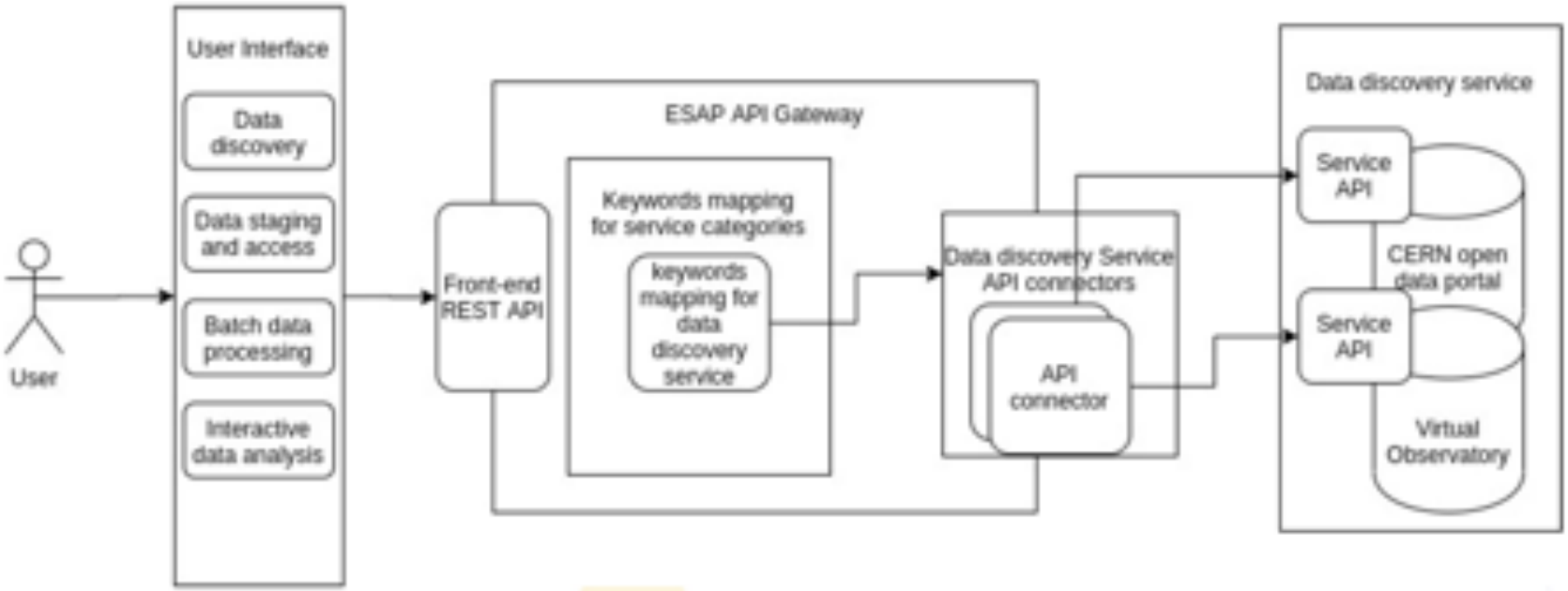
ESCAPE

ESAP Architectural Design



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

ESAP Architectural Design: Access to Virtual Observatory



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

EOSC next steps

European Commission

WP4 Next Steps

- Technology Forum results – update of the work plan
- Compile first results into Deliverable reports:
  - **D4.2** : **Intermediate report on the use of IVOA standards** – March 2020
  - **D4.4** : **Intermediate analysis report of VO data and service integration into EOSC** July 2020
- IVOA Interoperability meeting – 4-8 May [illegible] Cancelled COVID-19 ☹
- 1st Science with Interoperable Data School – May 26-28, 2020 Cancelled COVID-19 ☹ **(D4.3)**
- **Workshop for data providers in 2021**
- **VO Schools – to be re-scheduled**

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

ESCAPE information

- Access to information to be improved….

- ESCAPE pages, newsletter, communications:

https://www.projectescape.eu

- WP4 Technology Forum:
  https://indico.in2p3.fr/event/20005/

- ESCAPE Progress Meeting presentations:
  https://indico.in2p3.fr/event/20203/

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