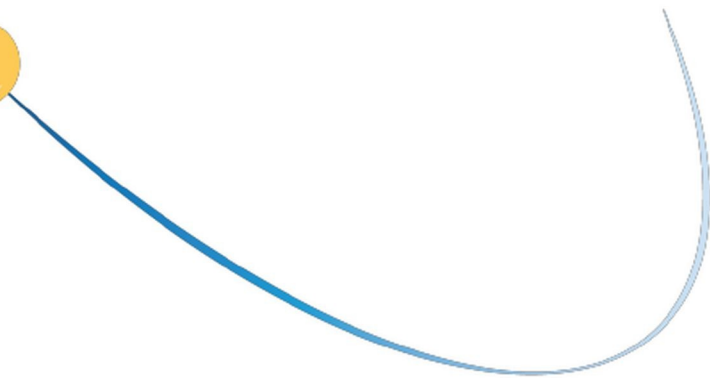




ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures





ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

CONNECTING ESFRIs to EOSC

Catherine Boisson, LUTH, Observatoire de Paris

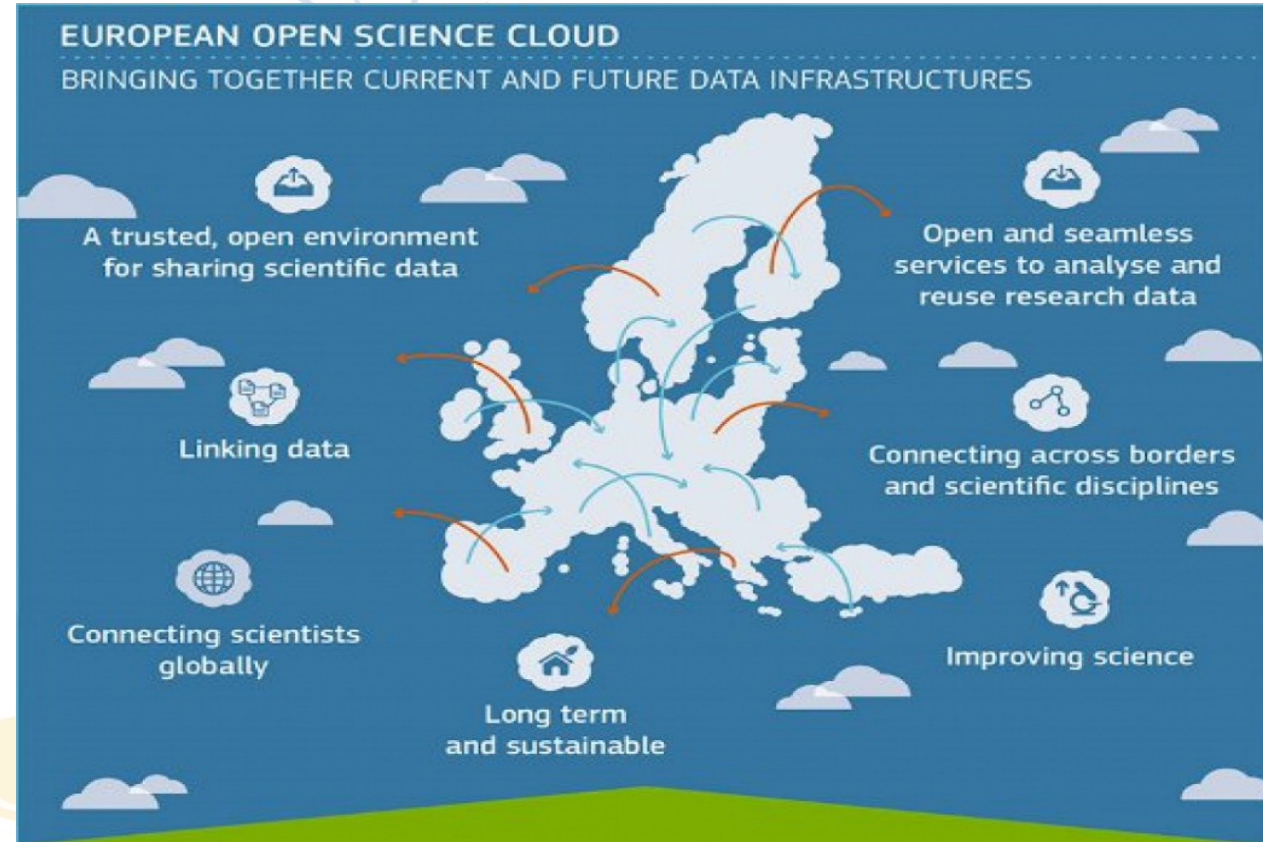
Own selection of essentials borrowed from many ESCAPE members presentations



- *EOSC is a cloud for research data in Europe that allows universal access to data*
 - *FAIR data and services for data storage, management, analysis and re-use across borders and disciplines*
 - *Added values for data-driven science, reproducible science, digital innovation*

H2020-INFRAEOSC-04-2018 call

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



A cluster action of Big-Science ESFRI RIs for setting up EOSC, implies technical and policy challenges.

as per the European Commission “EOSC Declaration”

- EOSC as a data infrastructure commons serving the need of scientists, providing functions delegated to community level, federating resources.
- Researchers should contribute to define the main common functionalities needed by their own community.
- A continuous dialogue to build trust and agreements among funders, scientists and service providers is necessary for sustainability.
- Data Sharing and Data Stewardship are the main issues.



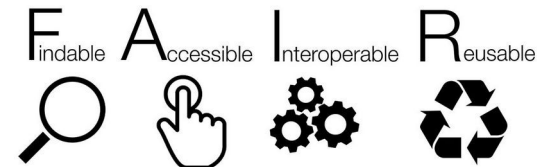
H2020-INFRAEOSC-04-2018 call

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.*

It is all about making data FAIR ...



- Allows to name essential characteristics of data sharing
- In use in some disciplines for a long time

FAIR ≠ Open, but the FAIR principles are everywhere in the Open Science context



Findable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

Accessible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

Interoperable

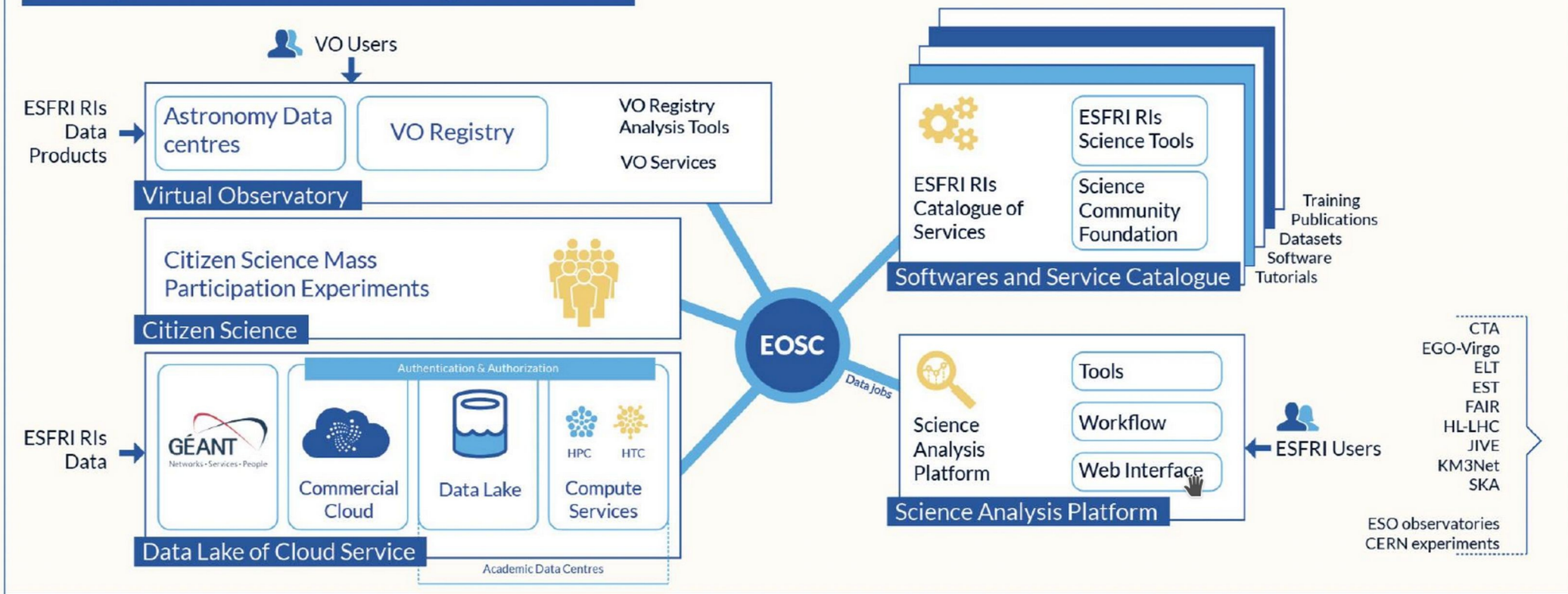
(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

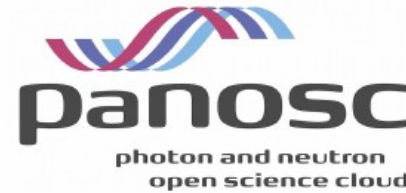
Reusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

What ESCAPE will deliver to EOSC ?

Management Innovation Networking Dissemination





- 5 Science Clusters and the European e-infrastructures are committed to working together ENVRI-FAIR, EOSC-LIFE, ESCAPE, PaNOSC, SSHOC
- Brings together 72 world-class RIs from the ESFRI roadmap and beyond to work on FAIR data management and connecting their communities to the EOSC
- The European e-infrastructure EDI, EUDAT, GEANT and OpenAIRE provide inclusive services.



- ESCAPE is based on the capacity building of the H2020 ASTERICS cluster of ESFRI projects (in astrophysics and astroparticle physics) addressing Big Data challenges and already succeeding in:
 - Enabling interoperability between the facilities,
 - Minimising fragmentation,
 - Encouraging cross-fertilisation and
 - Developing joint multi-messenger capabilities



*Astronomy ESFRI & Research Infrastructure Cluster
ASTERICS - 653477*



ESCAPE convenes a larger scientific community and a larger number of ESFRI projects concerned by Fundamental Science research

- 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:
 - the astronomy Virtual Observatory infrastructure
 - long-standing expertise of the particle physics community in large-scale distributed computing and big-data management.



ESCAPE convenes a large scientific community

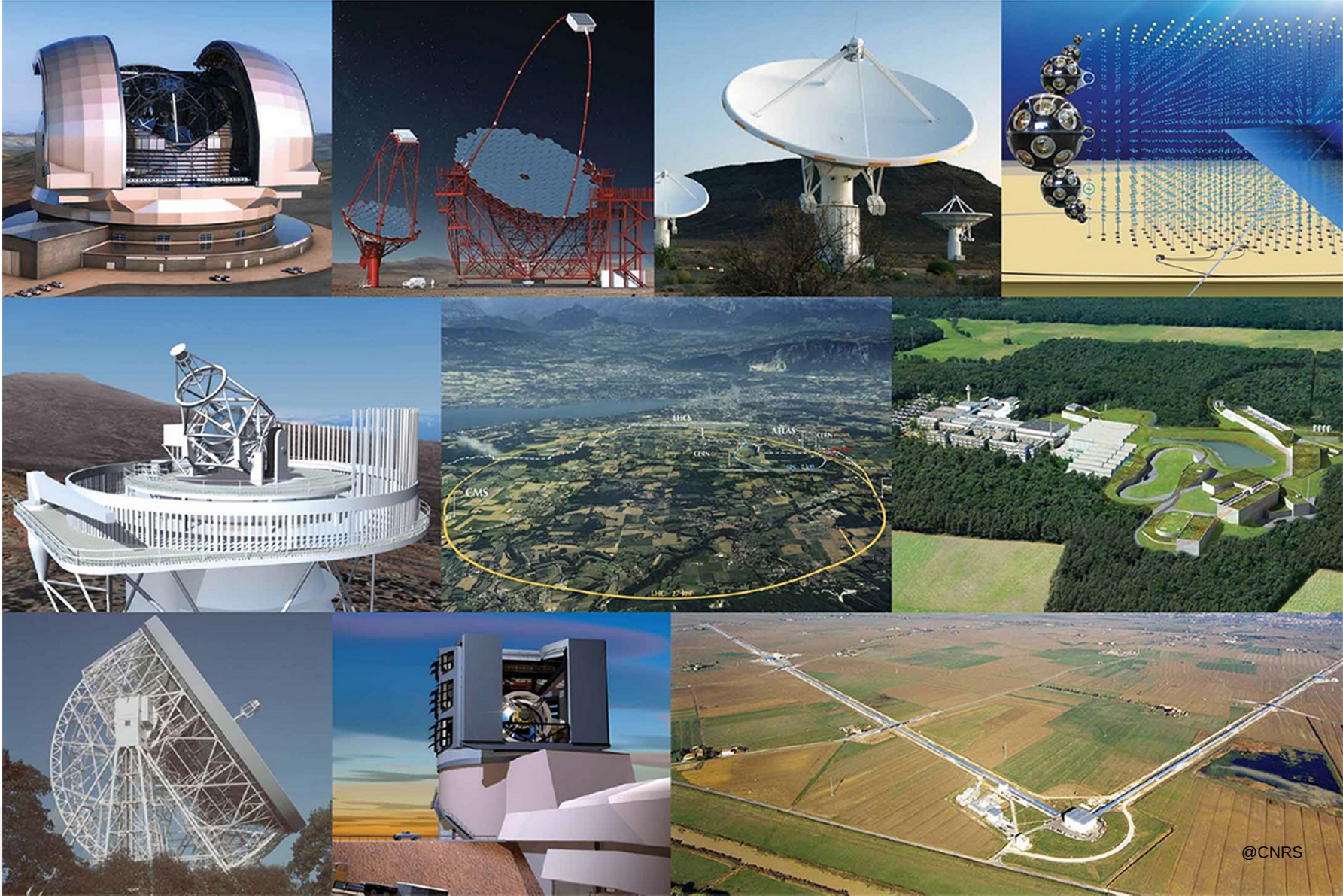
- 31 partners (including 2 SMEs)
- 7 ESFRI projects & landmarks: CTA, ELT, EST, FAIR, HL-LHC, KM3NeT, SKA
- 2 pan-European International Organizations: CERN, ESO (with their world-class established infrastructures, experiments and observatories).
- 4 supporting ERA-NET initiatives: HEP (CERN), NuPECC, ASTRONET, APPEC
- 1 involved initiative/infrastructure: EURO-VO
- 2 European research infrastructures: EGO and JIV-ERIC
- Budget: **15.98 M€**
- Started: **1/2/2019**
- Duration: **42 months** (end date 31/7/2022)
- Coordinator: **CNRS – G. Lamanna, LAPP**



Depuis 80 ans, nos connaissances
bâtissent de nouveaux mondes



ELT
CTA
SKA
KM3NeT
EST
HL-LHC
FAIR
JIV-ERIC
LSST
EGO-VIRGO





MAX-PLANCK-GESELLSCHAFT



JIVE
Joint Institute for VLBI
ERIC



THE UNIVERSITY
of EDINBURGH



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



Heidelberg Institute for
Theoretical Studies



CSIC
Spanish Council of Research



OROBIX
TOOLS WITH A MIND



rijksuniversiteit
 groningen



Royal Observatory
of Belgium



Trust-IT Services
Communicating ICT to markets



ELT
CTA
SKA
KM3NeT
EST
HL-LHC
FAIR
JIV-ERIC
LSST
EGO-VIRGO

ESCAPE ESFRI facilities aligned expectations

- Big-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 contours of current state-of-the-art.
- “Observatory” and “Facilities” type of operation requires global open access and long-term sustainability of the extremely large volume of FAIR research data and services of the ESFRI facilities.
- Training and extension of FAIRness standards and tools for data access and data preservation.
- Operating a common open innovation environment.
- Already existing inter-RI cross-talk, intersections; overlapping competence and authority of national stakeholders.



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 exposed 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 SMEs and society in knowledge discovery.



“Management, Innovation, Networking & Dissemination”

Resp: G. Lamanna, LAPP

- Visibly promote ESCAPE to the global, European and national communities by attending relevant meetings, international conference, including symposia on the EOSC implementation and making high quality presentations.
- Represent and promote ESCAPE in interface with strategy and policy bodies such as the EC, ESFRI committees, EOSC-hub strategy committee, e-IRG, Open Science Policy Platform, RDA and the High-Level Expert Group on EOSC.
- Coordinate the implementation of the “**ESCAPE Test Science Projects**” proposed to validate ESCAPE services and Open Science at the end of the project (Dark Matter; Extreme Universe/MM).

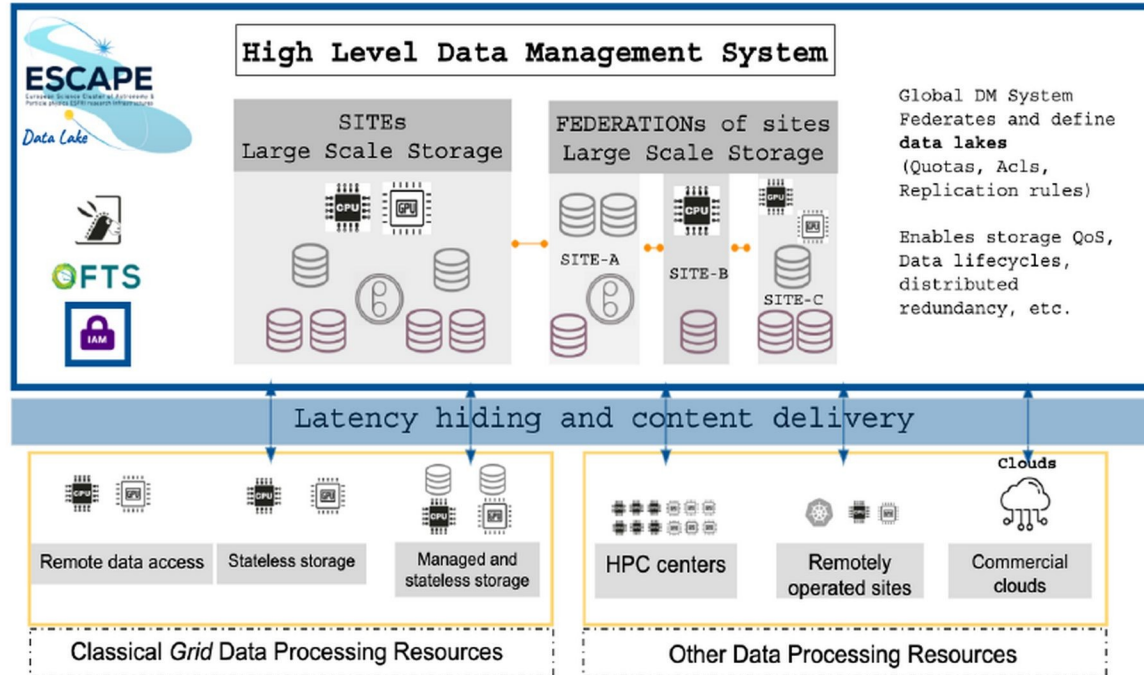


“Data Infrastructure for Open Science”

Resp: S. Campana, CERN

- Contribute to the federation of global EOSC resources through an implementation of the **data-lake** concept to manage extremely large volumes of data up to the multi-exabyte scale
- Design, implement and operate a cloud of data services for open access and open science at Exabyte scale
- The backbone of the Data Lake are the well experienced large national data centers





- Leaves to the science projects the flexibility to choose the services and layout most suitable to their needs. Provides **global data management** orchestration.
- Contribute to deliver **Open Access and FAIR data services**: relies on trustable data repositories; enables data management policies; hides the complexity of the underlying infrastructure providing a transparent data access layer



“Open-source Scientific software and Service Repository”

Resp K. Graf, FAU

- Support for “Scientific software” as a major component of the ESFRI-RI “data” to be stored in EOSC
- Establish a **community-foundation** of EOSC-ready open-source software and services (innovative workflows, common libraries, software development...)
- Expose/share software to users via the **EOSC catalogue** of services
- Train and guide the scientists/users



“Connecting ESFRI projects to EOSC through VO framework”

Resp M. Allen, CDS-CNRS

- Extend FAIR standards, methods, tools of the Virtual Observatory to a broader context
- Demonstrate EOSC’s ability to include existing platforms
 - **Integrating IVOA architecture** into EOSC
 - Promoting the existing VO infrastructure
 - Metadata standards - vocabularies, units, data models
 - Technical standards - data formats, service protocols
 - High level standards - data discovery, data access



“Connecting ESFRI projects to EOSC through VO framework”

- Integrating IVOA architecture into EOSC
 - Updating and improving the VO infrastructure
 - Some ESFRIs are in areas well represented by the VO
 - Some ESFRIs are in areas new to the VO.
 - New requirements, new workflows, new data structures
 - Common challenges
 - cloud compute data storage, data transfer etc
 - authentication, authorization, identity



“European Scientific Analysis Platform”

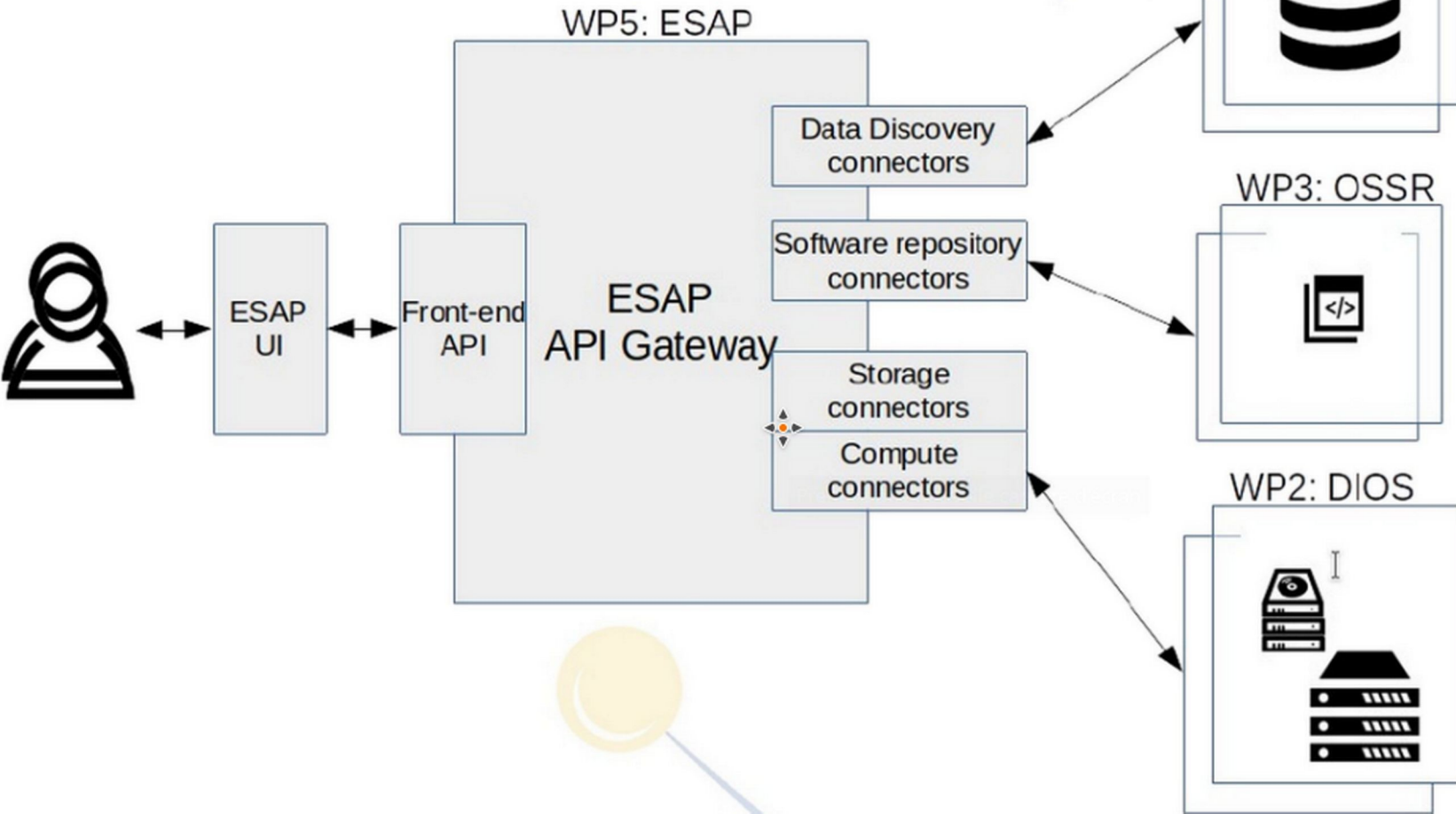
Resp. M. Van Haarlem, ASTRON-NWO

- Implement **Flexible and Expandable Science Platform** to make EOSC a working interface and bring analysis to data
- Support users to:
 - identify & stage existing data collections
 - tap into software tools & packages developed by ESFRIs, bring own custom workflows
- Focus on core common functions to support more communities
 - flexibility rather than single platform for all users



ESAP Architectural Design

WP5: ESAP



“Engagement and Communication”

Resp. S. Serjeant, Oxford Open University

- Towards wider public – Citizen Science
 - trains and educates the community in the usage and implementation of the ESCAPE services and ESFRI facilities, in line with the FAIR principles.
 - engages the society at large to foster innovation in science and technology, contribute to real scientific discoveries and support the implementation of EOSC via the next generation of university students, scientists and engineers, who are the future users of the ESFRI facilities.



Data Lake:

- Build a scalable, federated, data infrastructure as the basis of open science for the ESFRI projects within ESCAPE. Enable connection to compute and storage resources.



Software Repository:

- Repository of "scientific software" as a major component of the "data" to be curated in EOSC. Implementation of a community-based approach for the continuous development of shared software and for training of researchers and data scientists.



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.



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



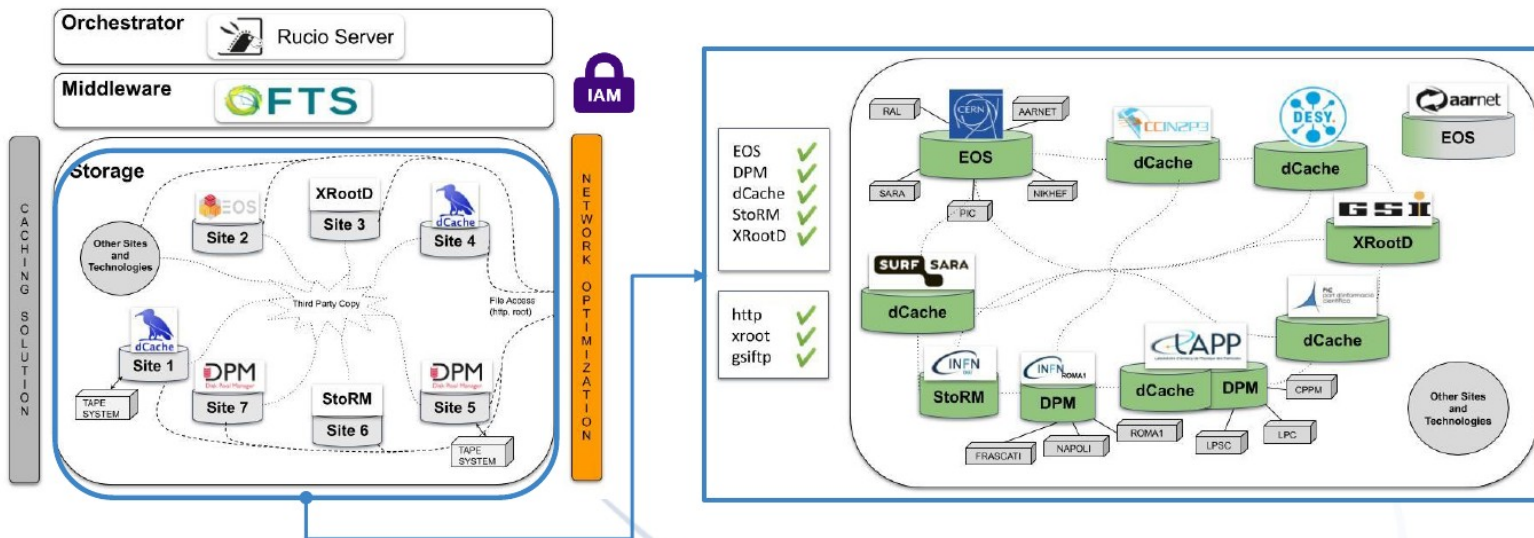
The logo features a blue arc at the top left that curves towards a blue starburst. At the bottom, a yellow circle is connected to a blue arc that curves towards the right.

ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

First achievements : a functional Data Lake pilot

- ⑩ Pilot Data Lake with **10 storage endpoints functional**:
CERN, DESY, GSI, IFAE-PIC, IN2P3-CC, INFN-CNAF, -ROMA, -Napoli, LAPP-MUST and SURF-SARA
- ⑩ The high level Data Lake orchestration layer is **consolidated**



First achievements : Science in the Data Lake

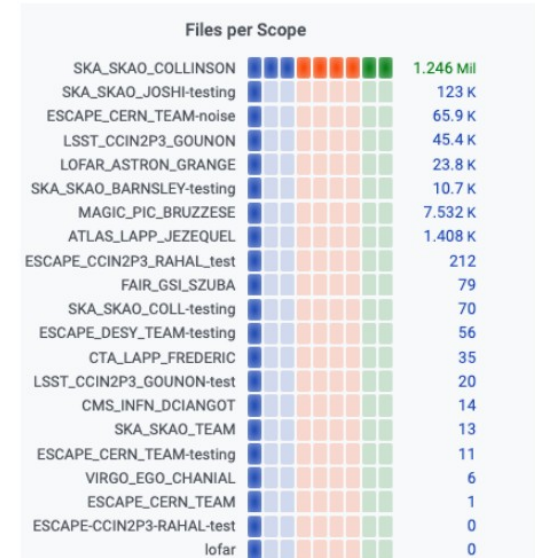
⑩ Strong involvement of ESFRI RIs and other experiments:

- **Data injection** within the Data Lake by:

ATLAS, CMS, CTA, FAIR, LOFAR, LSST, MAGIC, SKA, and VIRGO/EGO

⑩ Data management demonstrator from **Astroparticle, Radio-astronomy, Gravitational Waves, Cosmology** and **Particle Physics** communities together on a **common** data management infrastructure

⑩ Pipeline data analysis tests currently in progress



ESCAPE OSSR and Development Platform – how to ease the publication and integration process?

From a single
click

- Publishes source code
(updates your existing record with
new versions)



- Long term archived
- Findable
- Citable



1. Make a new tag (release)
2. Let the CI do the rest

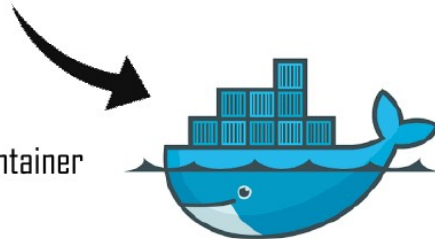
- builds a
singularity image



- publishes
singularity image

under
dev

- builds a
docker container



- publishes

