FAIR in the EOSC Association

Sarah Jones, GÉANT
sarah.jones@geant.org
Twitter: @sarahroams

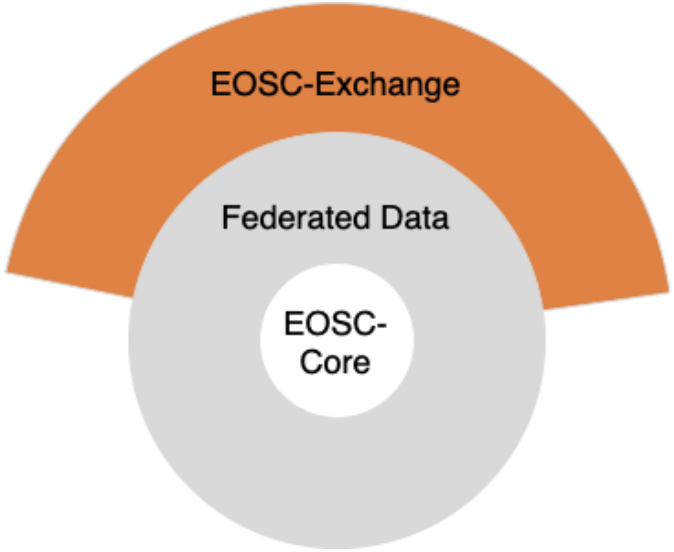
DeiC conference

3 November 2021

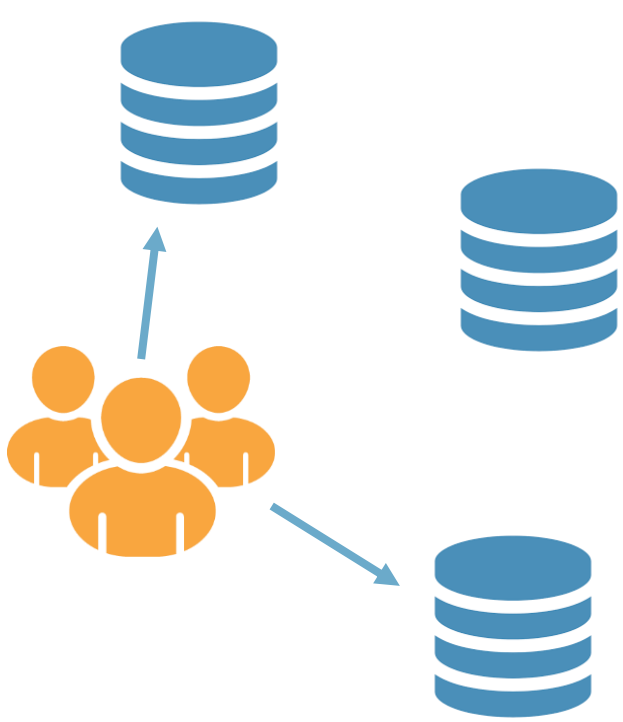
EUROPEAN OPEN SCIENCE CLOUD

What is EOSC?

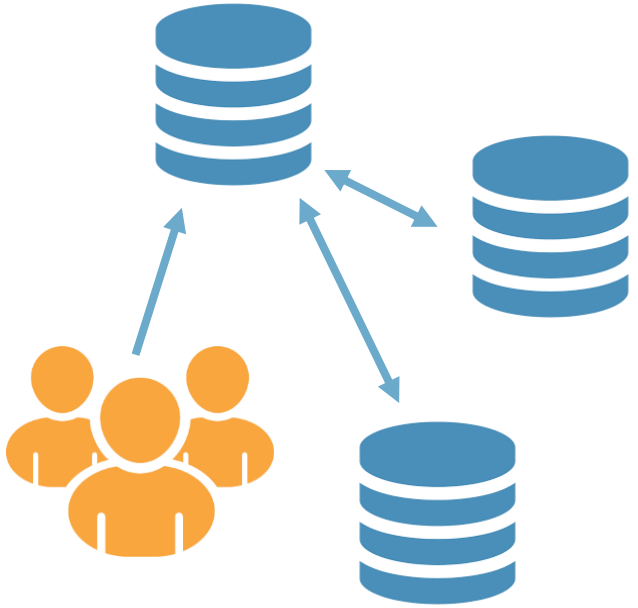
- A web of FAIR data and services
- Federation of eInfra and research infrastructures
- Environment in which data can be brought together with services to perform analyses and address societal challenges



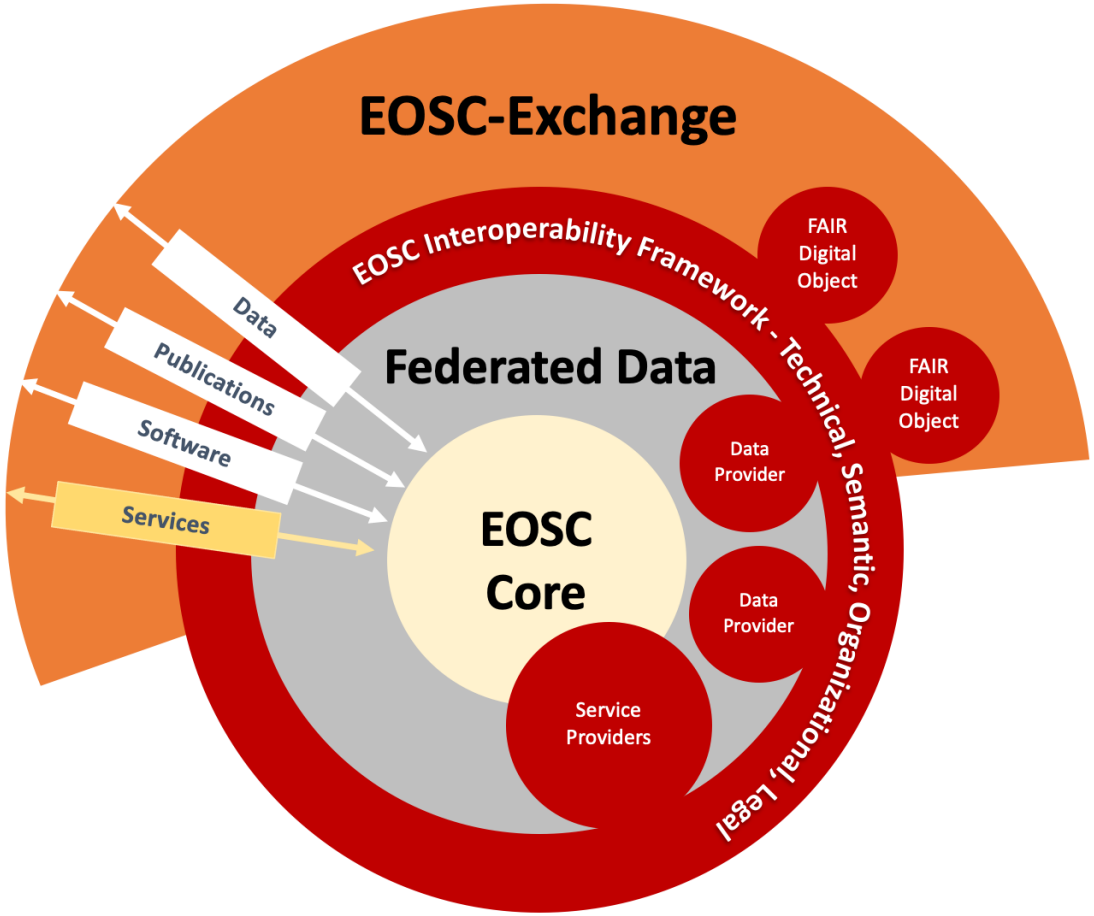
Aims to enable multidisciplinary discovery & use



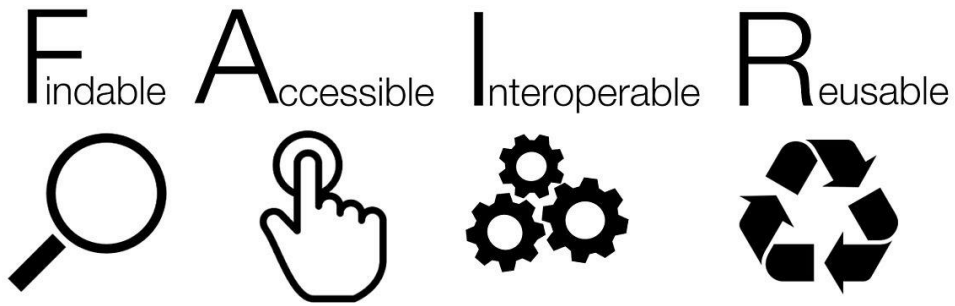
Disconnected silos to a federated infrastructure providing added value to researchers



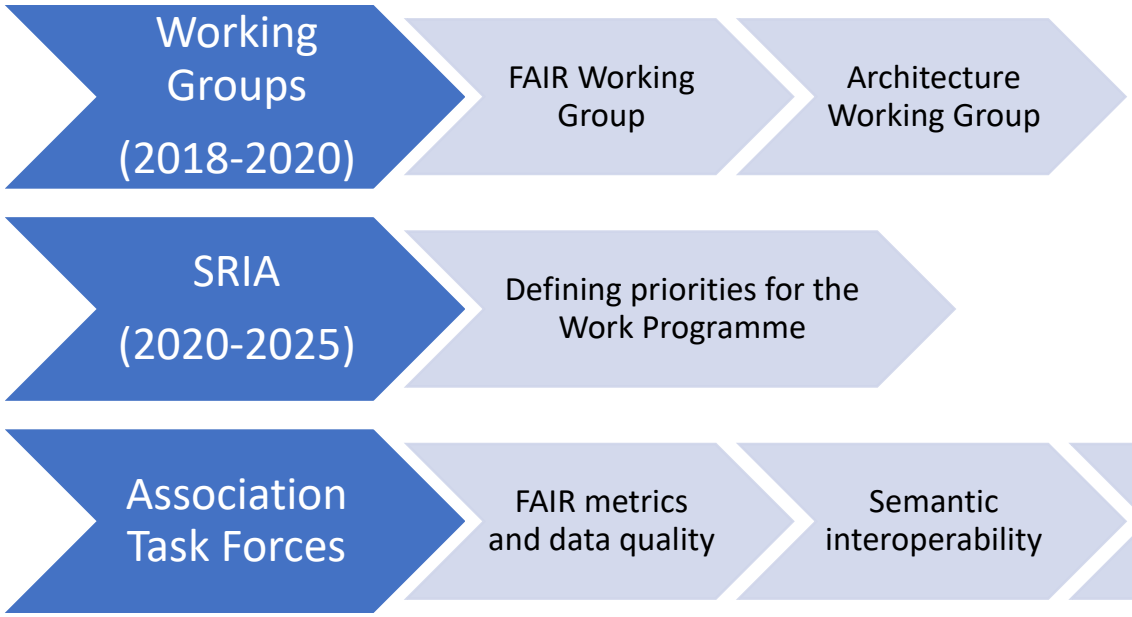
FAIR is central to principles in EOSC



- Is the glue that connects data & services
- Requirement for FAIR to support reuse
- Use community standards
- Share all types of output (openly)



Timeline of implementing FAIR in EOSC



FAIRSFAIR
Fostering Fair Data Practices in Europe

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

ENVRI FAIR

SSHOC
social sciences & humanities open cloud

EOSC-Life

panosc
photon and neutron open science cloud

EOSC Future

Recommendations via governance working groups / task forces and implementation via projects and infrastructures

EOSC FAIR Working Group: Task Force remit

Task Force Leads
-Rob Hooft
-Marta Teperek

Six recommendations report:
https://doi.org/10.2777/986252

FAIR practice

PID policy

Task Force Leads
-Peter Wittenberg
-Rachael Kotarski

EOSC PID policy
https://doi.org/10.2777/926037

Task Force Leads
-Oscar Corcho
-Krzys Kurowski

EOSC interoperability framework
https://doi.org/10.2777/620649

Interoperability

Metrics & certification

Task Force Leads
-Francoise Genova
-Magnus Aronsen

EOSC metrics: https://doi.org/10.2777/70791
FAIR services: https://doi.org/10.2777/127253

EUROPEAN OPEN SCIENCE CLOUD

FAIR practice report

divides not necessarily disciplinary - often depend on data types, country/region, age… If there is a **culture** and support within the community then FAIR practices develop

Six recommendations for Implementation of FAIR practice

https://doi.org/10.2777/986252

	EOSC	Research funders	Institutions	Policy-makers	Coordination fora	Standards bodies	Data service providers	Publishers
1. Fund awareness-raising, training, education and community-specific support	√	√	√					
2. Fund development, adoption and maintenance of community standards, tools and infrastructure	√	√			√	√	√	
3. Incentivise development of community governance	√	√			√			
4. Translate FAIR guidelines for other digital objects	√	√		√	√	√		
5. Reward and recognise improvements of FAIR practice	√	√	√	√				
6. Develop and monitor adequate policies for FAIR data and research objects	√	√	√	√				√

PID policy for EOSC

A Persistent Identifier (PID) Policy for EOSC

https://doi.org/10.2777/926037

- Context
- Principles
- Generic PID definitions
- Roles and responsibilities
- PID applications
- PID types
- PID services and PID service providers
- Governance and sustainability

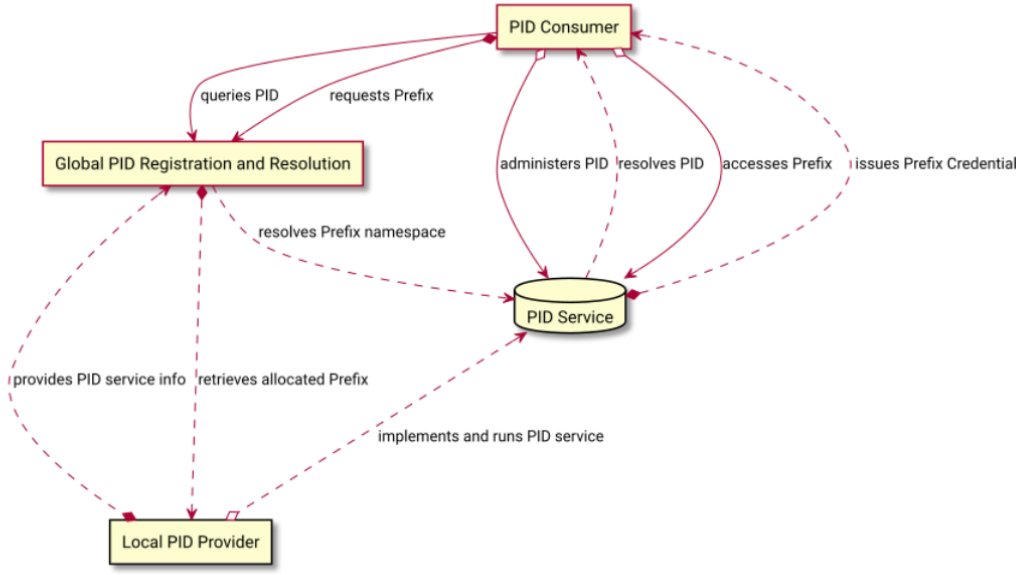
"The EOSC PID Policy identifies that an **ecosystem of PID infrastructures is needed** to support a wide variety of scientific applications and offer sufficient flexibility and capacity. The ARDC PID portfolio also aims to cover off important elements of the research enterprise (**people, organisations, data, software, publications, instruments,** etc). We are enthusiastic to see the outcomes of the EOSC PID Policy and interested in collaboration wherever that may be useful for all parties." ARDC

PID Architecture

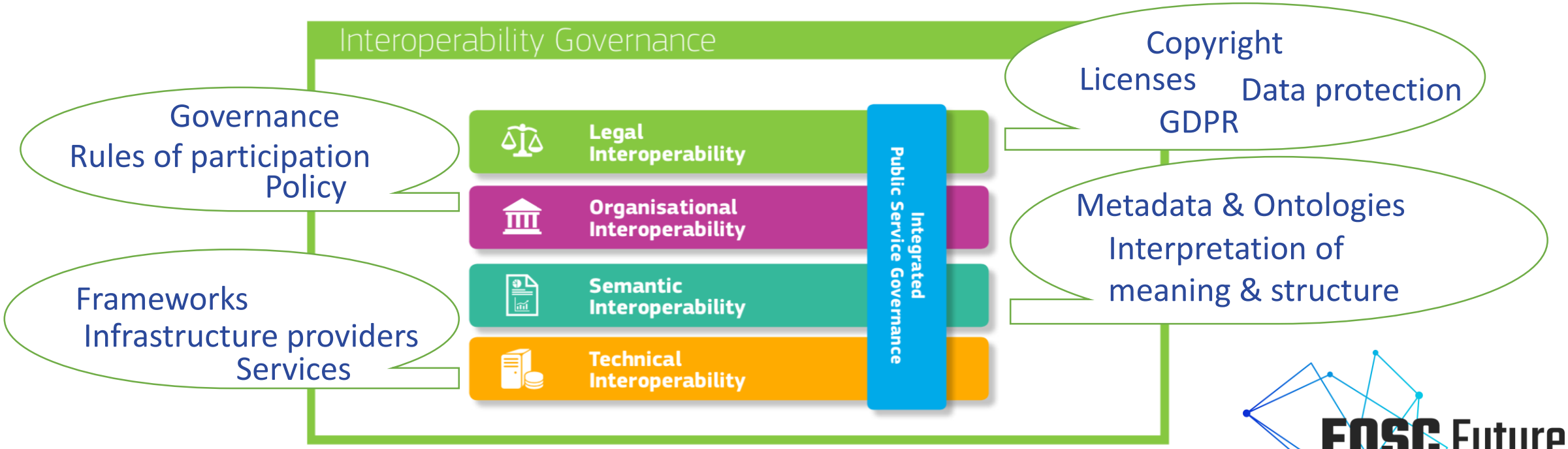
Three highest ranked priorities:

- Calls for a governance structure, as the PID domain is lacking an overarching organisation in which the different PID frameworks come together e.g. like ICANN for DNS
- Calls for a global service that enables the resolution of any kind of PID. Suggests the Handle system as a starting point.
- Support for emerging PID types to standardise kernel information e.g. for instruments and organisations

https://doi.org/10.2777/525581



EOSC Interoperability Framework



The European Interoperability Framework four levels of interoperability

https://doi.org/10.2777/620649

EOSC Future

Being taken forward through EOSC Future

FAIR metrics for EOSC

- Metrics recommend a subset of the c.40+ RDA FAIR Data Maturity Model metrics
- Phase what is required to allow all communities to engage
- Metrics and associated tools to be thoroughly tested

Aspect	2021	2024	2028
Discovery	• Metadata is provided to allow discovery • Metadata includes the identifier for the data	• Metadata is offered in such a way that it can be harvested and indexed	• Metadata is guaranteed to remain available, after data is no longer available
Licence	• Metadata includes information about the license under which the data can be reused	• Metadata refers to a standard reuse license	• Metadata refers to a machine-understandable reuse licence
Standards		• Data/Metadata complies with a community standard	• Data/Metadata is expressed in compliance with a machine understandable community standard • Metadata use FAIR-compliant vocabularies.

FAIR service certification

- Recommends certification for repositories but **certification status cannot be a necessary condition to be included in EOSC yet**
- Certification schema should be aligned with FAIR (CTS doing this)
- Should establish criteria to certify other elements e.g. PID systems and vocabulary / metadata registries

https://doi.org/10.2777/127253

Strategic Research and Innovation Agenda

- FAIR is a guiding principle, informing infrastructure implementation
- Key Action Areas and priorities have been set on:
  - AAI
  - Identifiers
  - Metadata and Ontologies
  - FAIR metrics and certification
  - EOSC Interoperability Framework

https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

Key areas covered in present work programme

HORIZON-INFRA-2021-EOSC-01-05

- Uptake of and compliance with FAIR data principles and practices
- Support research communities to implement existing or emerging metrics and make use of the FAIR data maturity model
- Foster alignments with existing interoperability frameworks and standards

HORIZON-INFRA-2022-EOSC-01-04

- Alignment of global standards and methodologies for FAIR

HORIZON-INFRA-2021-EOSC-01-03

- A PID meta resolver, infrastructure and PID graph
- Federated search, schemas, APIs and crosswalks for improved discovery and interoperability

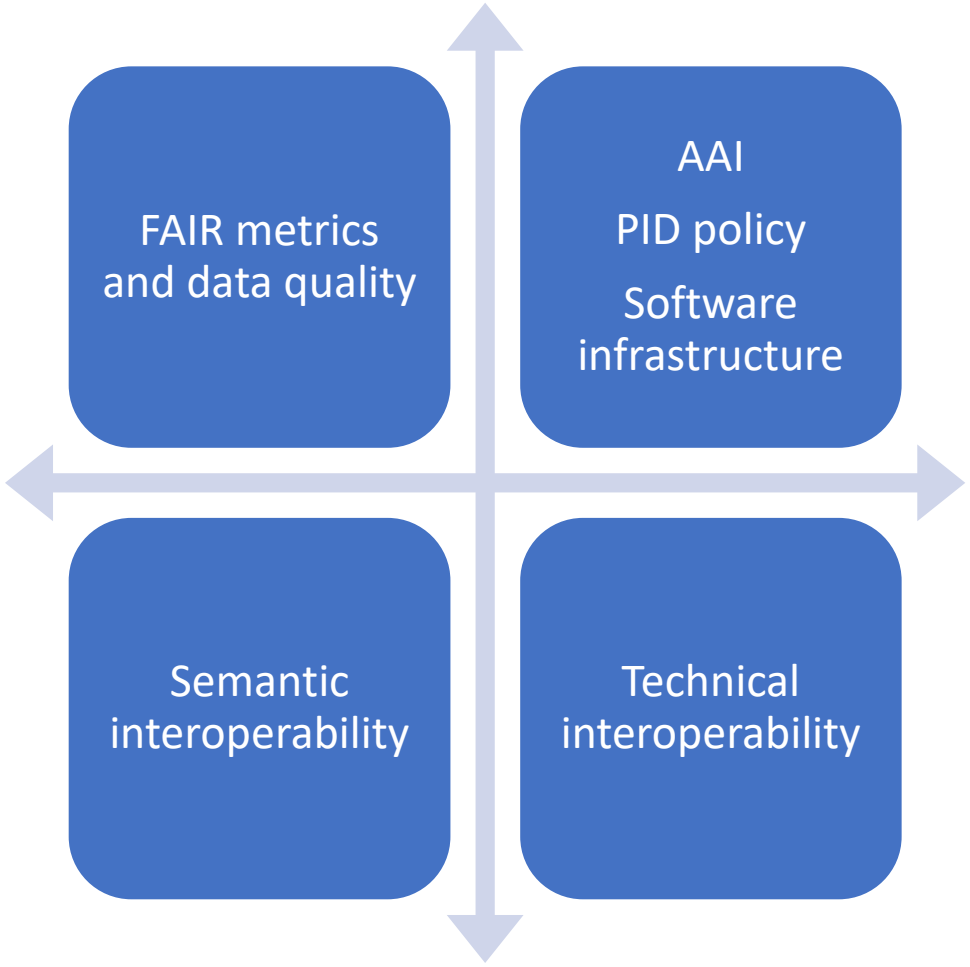
HORIZON-INFRA-2021-EOSC-01-02

- Effective business models

HORIZON-INFRA-2022-EOSC-01-01

- A rewards and recognition system based on a new generation of metrics and indicators

FAIR in the EOSC Association Task Forces



- Several Task Forces that are core to the FAIR agenda (i.e. metrics and interoperability)
- Many others implement core infrastructure and services needed to enable FAIR

https://www.eosc.eu/news/draft-charters-eosc-association-task-forces-published

FAIR in the EOSC Association Task Forces

- **FAIR metrics**: testing implementation, recommending tools, agreeing data quality measures…
- **Semantic interoperability**: minimal metadata, catalogue integration, crosswalks / mappings between standards, registries of semantic artefacts…
- **AAI:** common global ecosystem for identity and access control – AARC + gaps + governance
- **PID policy**: recommend types of identifiers, global PID resolution, infrastructure etc
- **Infrastructure for quality research software:** ensuring software is managed and recognised
- **Technical interoperability:** architecture for implementing Interoperability Framework

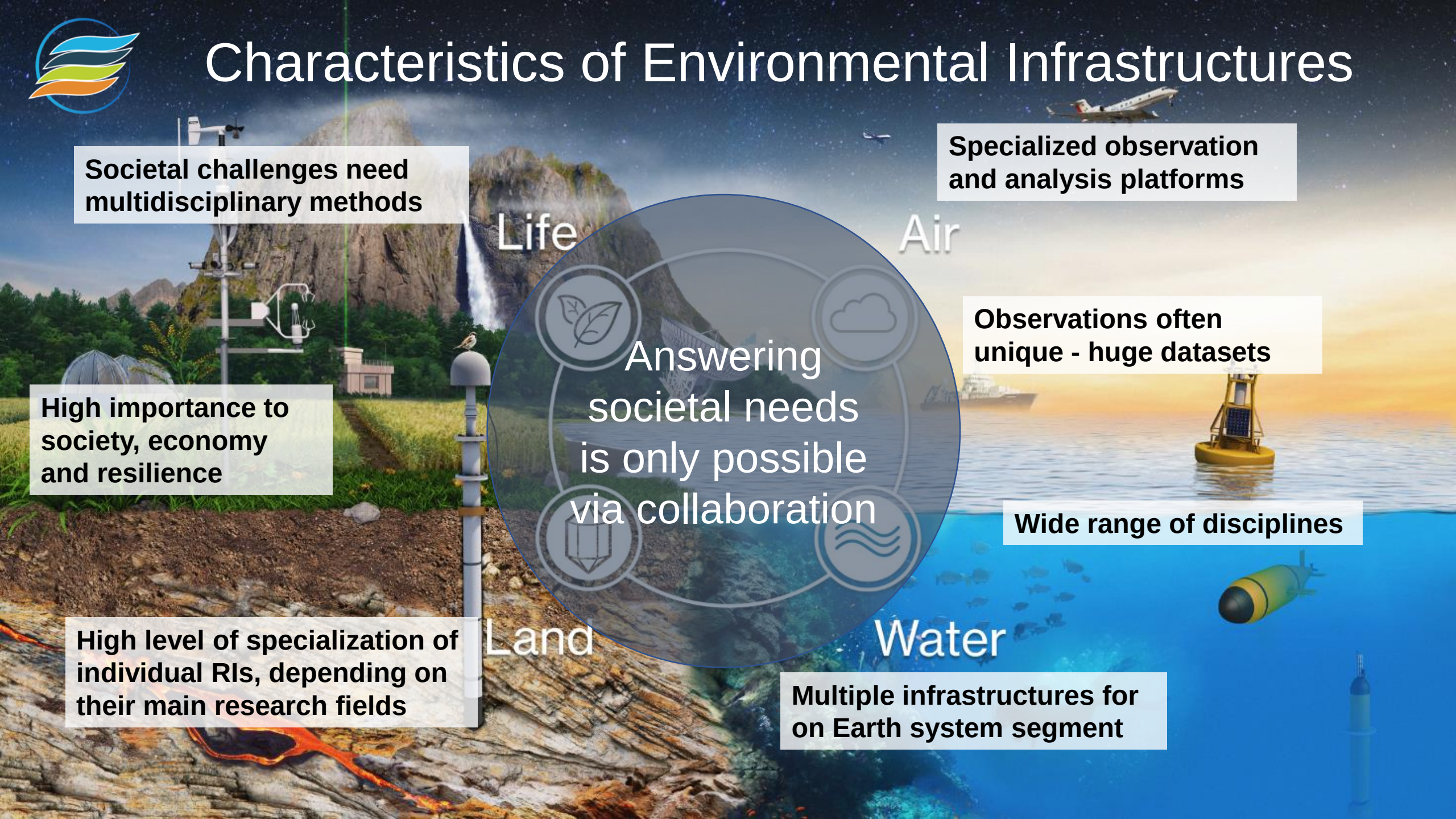
https://www.eosc.eu/news/draft-charters-eosc-association-task-forces-published

Practical implementation in projects

Image by Glenn Carstens-Peters https://unsplash.com/photos/npxXWgQ33ZQ

EUROPEAN OPEN SCIENCE CLOUD

Characteristics of Environmental Infrastructures



Societal challenges need multidisciplinary methods

Specialized observation and analysis platforms

Observations often unique - huge datasets

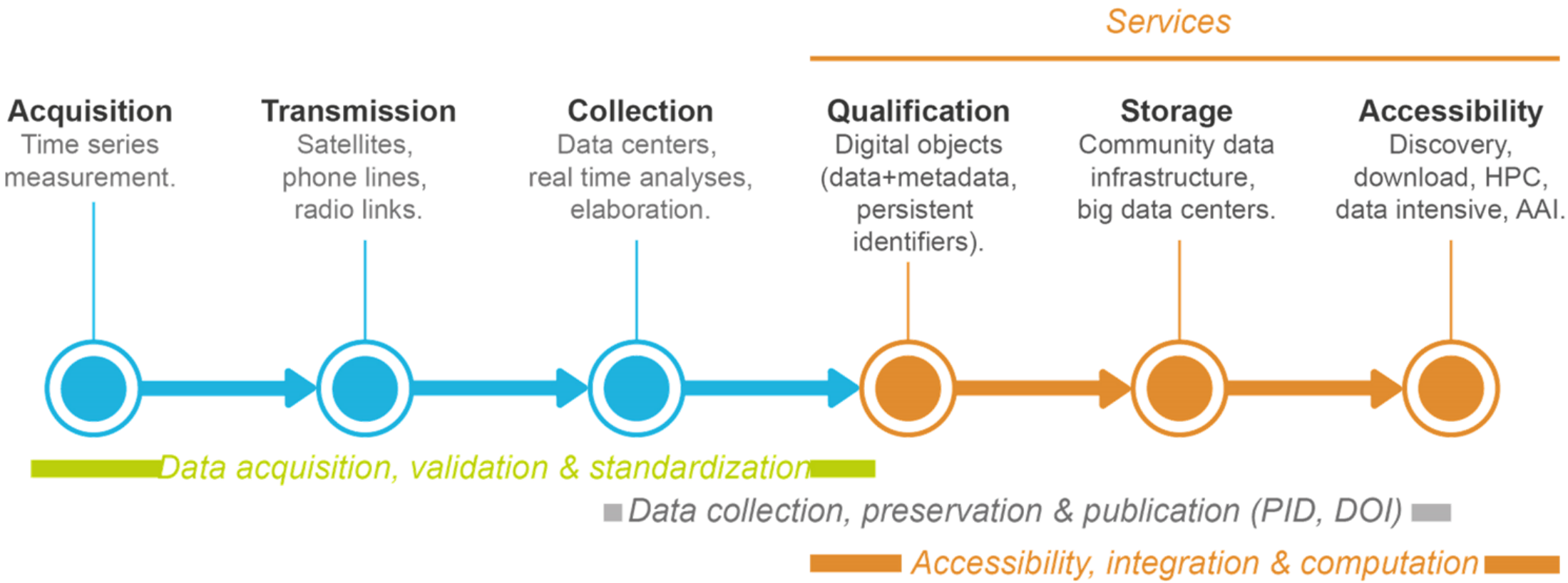
High importance to society, economy and resilience

Wide range of disciplines

High level of specialization of individual RIs, depending on their main research fields

Multiple infrastructures for on Earth system segment

Data Lifecyle Challenges in the ENVRI Domain



Standard Operating Procedures

MetaData formats
Naming conventions
Data analysis workflows

Use of PIDs
Workflow management
Provenance

Trusted repositories
Sustainable infrastructures

Vocabularies
Machine access
AAI
Licenses

FAIR Assessment Methods used in ENVRI-FAIR

FAIR
Assessment Questionnaire

ENVRI FAIR

GO FAIR

How to go FAIR

FAIR
Implementation Profile (FIP)

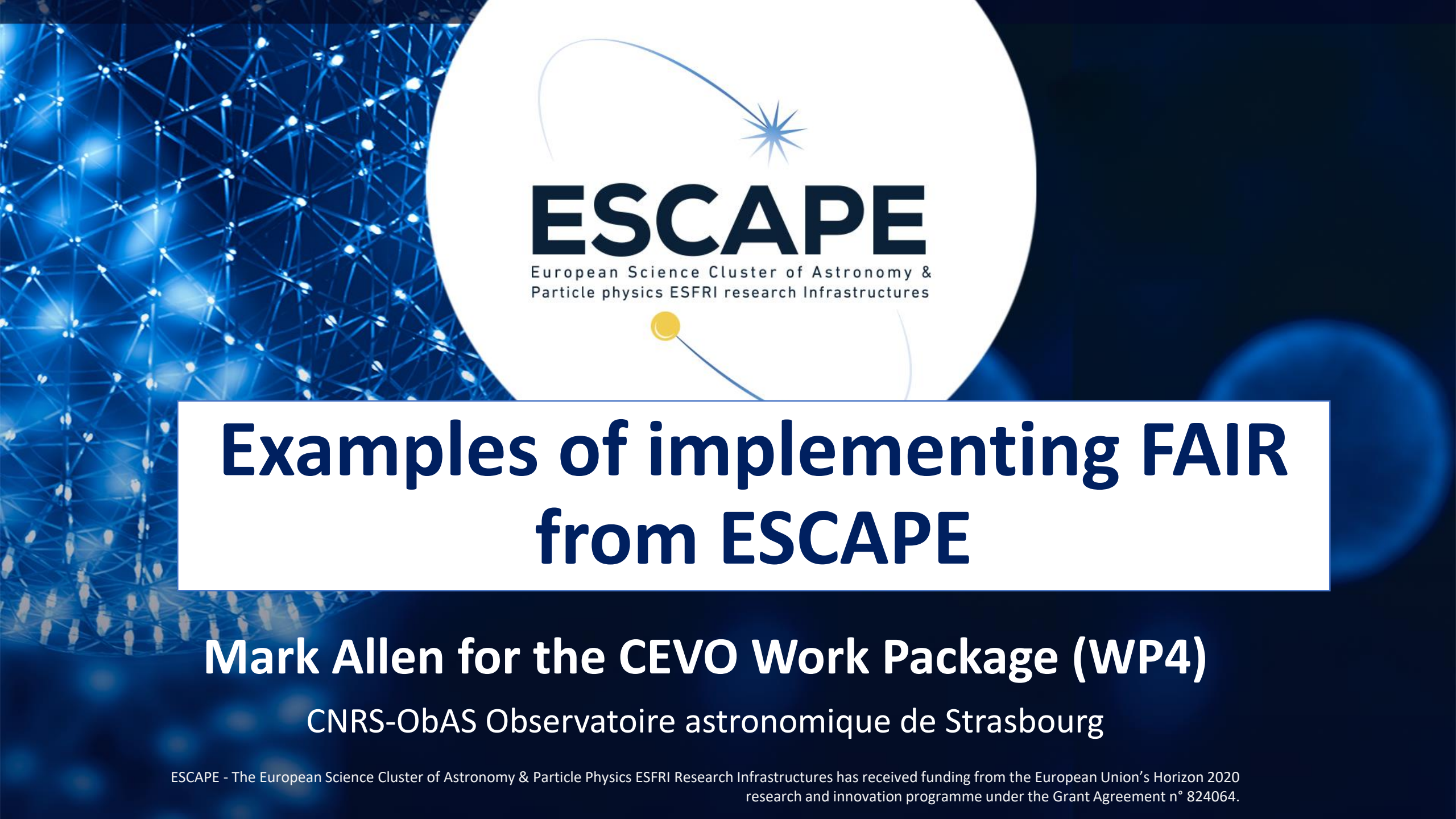
Community related metadata
F1: Identifier type, MD/D
F2: metadata schema; MD
F3: MD-D linking mechanism
F4: indexed search engines, MD/D
A1: communication protocol, MD/D
A1.1: authentication & authorisation technique, MD/D
A1.2: metadata longevity, MD
I1: knowledge language, MD/D
I2: annotation, encoding vocabularies: MD/D
I3: data schema, D,
R1.1: license, MD/D
R1.2: provenance model, D
R1.3: the FIP

GO FAIR Metadata 4 Machines

FAIR Implementation Profile

FAIR Data Point

ENVRI FAIR



Examples of implementing FAIR from ESCAPE

Mark Allen for the CEVO Work Package (WP4)

CNRS-ObAS Observatoire astronomique de Strasbourg

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.

FAIR concepts are integrated into all aspects of ESCAPE

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

Interoperability Standards
Metadata / Protocols
International context -

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.

Astronomy Virtual Observatory framework as part of EOSC

Integration of an operational interoperability framework for FAIR

Domain specific thematic services supporting Open Science

IVOA standards for implementation of FAIR

Brings Astronomy metadata standards into EOSC context

IVOA standards responding to the needs of ESFRI, RIs and researchers

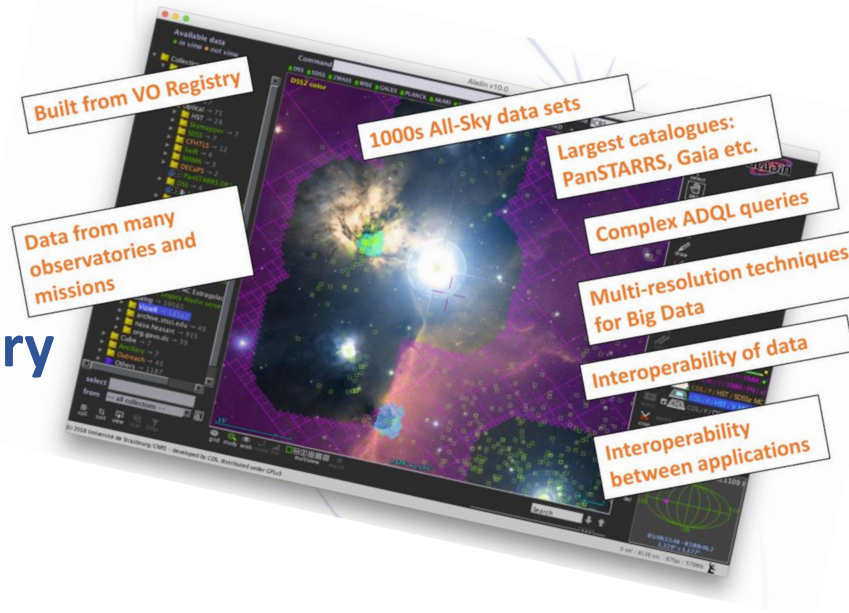
See Astronomy use case in SRIA, and EOSC Interoperability Framework

EOSC to enable next steps of the astronomical Virtual Obsevatory

Connection to computing and integration into ESCAPE platform

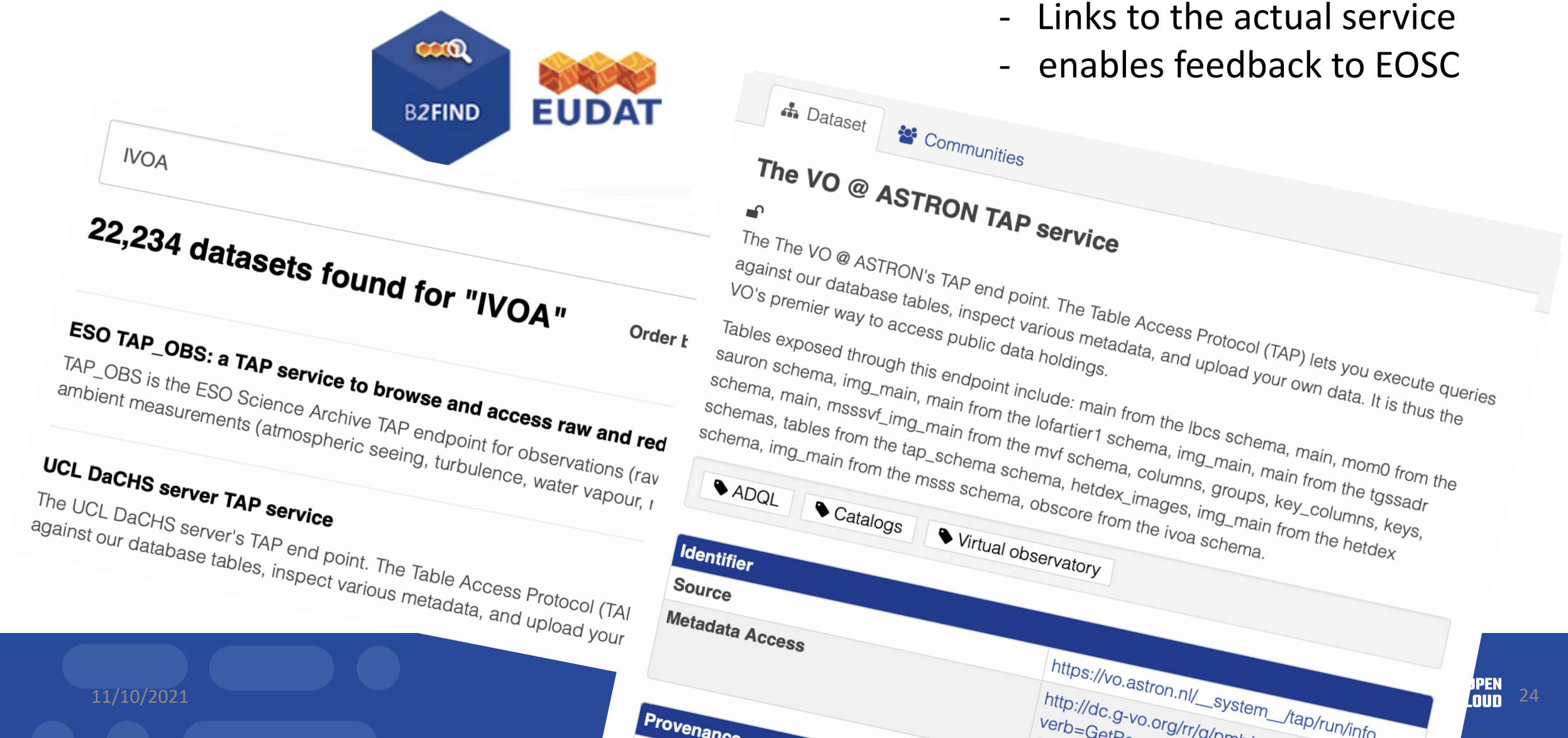
Scalability for big data

Data stewardship practices of Astrophysics in EOSC context



VO in B2FIND - a step to EOSC

- **Demonstrates 1st level of metadata compatibility**
  - Links to the actual service
  - enables feedback to EOSC



ESCAPE connecting to EOSC… a work in progress

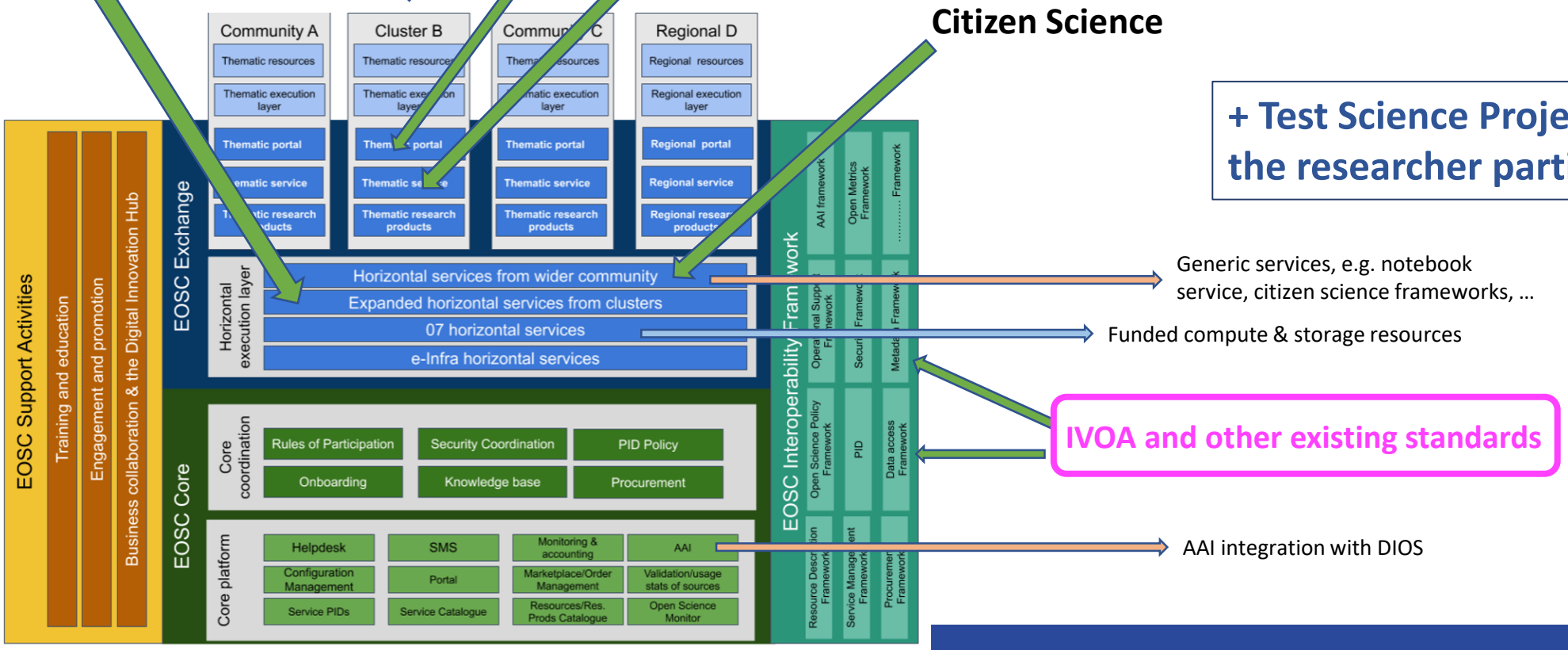
DIOS, OSSR
Data Lake
Software Repository

ESCAPE

ESAP Science Analysis Platform

Virtual Observatory services

Citizen Science



+ Test Science Projects to enhance the researcher participation in EOSC

Generic services, e.g. notebook service, citizen science frameworks, …

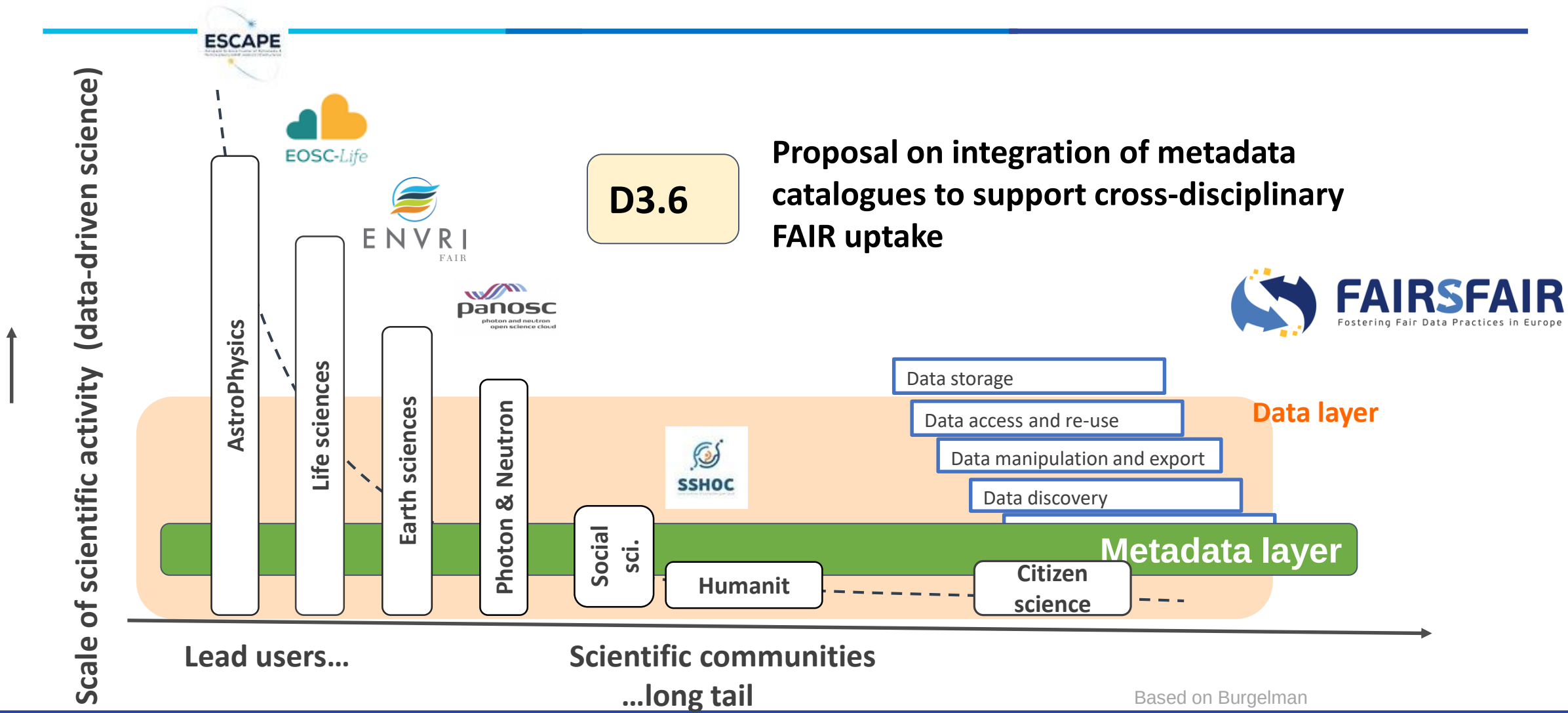
Funded compute & storage resources

IVOA and other existing standards

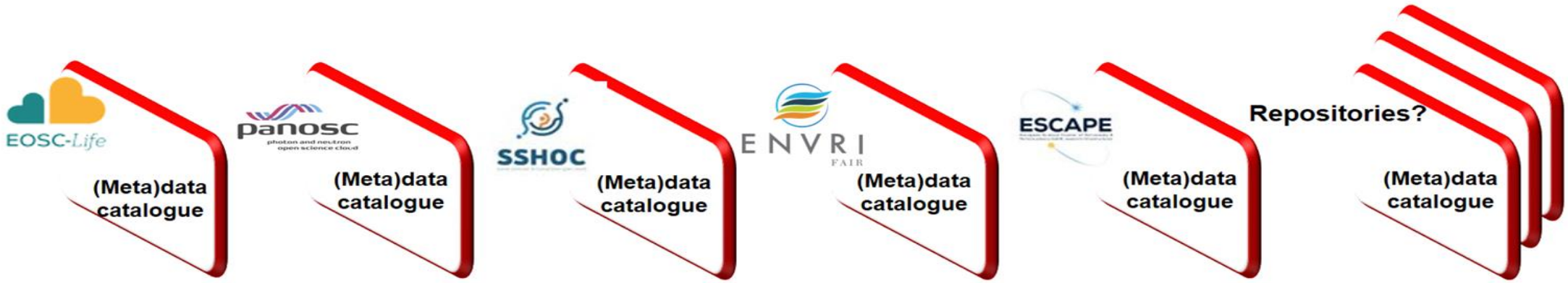
AAI integration with DIOS

EUROPEAN OPEN SCIENCE CLOUD

Metadata catalogue integration in FAIRsFAIR

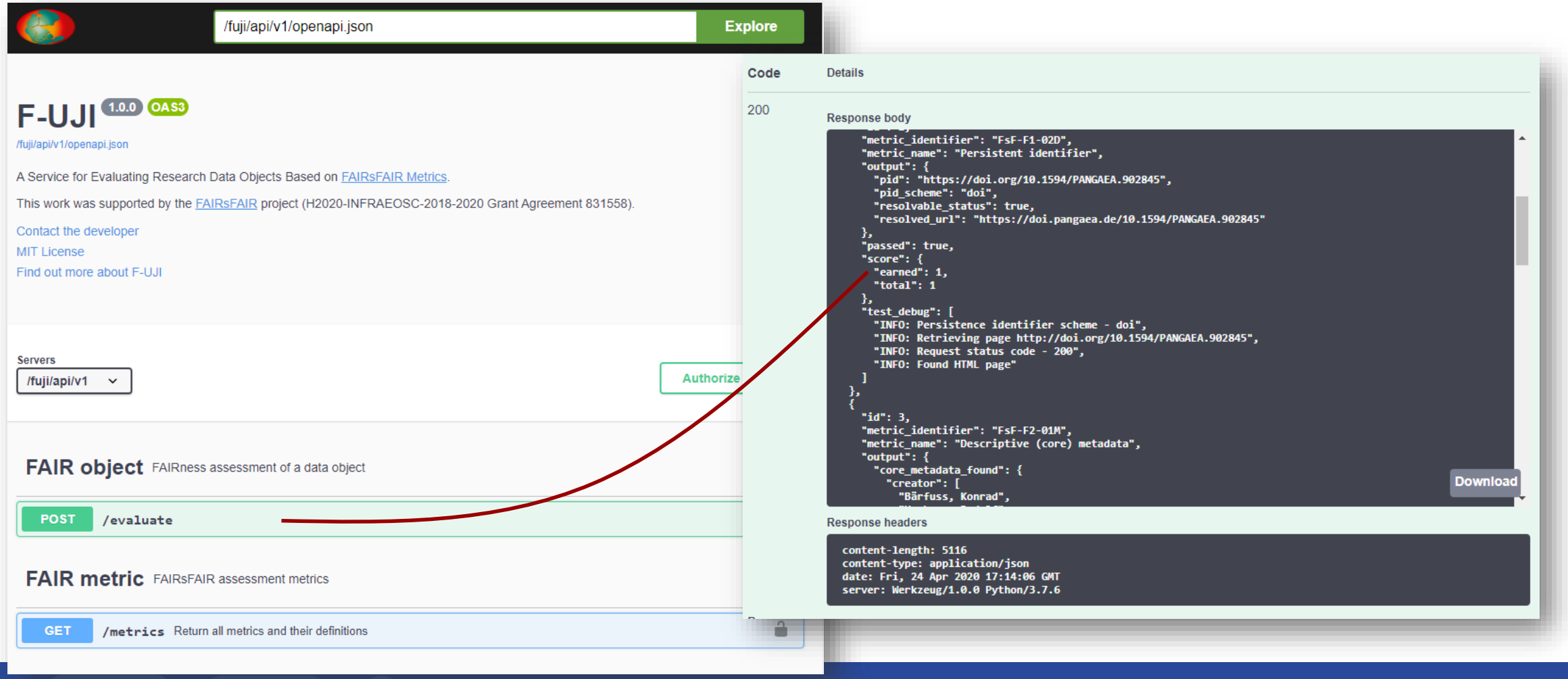


Identify domain agnostic metadata schema



B2FIND Find Research Data — **Integrated (Meta)data catalogue** — B2FIND Find Research Data

F-UJI – An Automated FAIR Data Assessment Tool



EUROPEAN OPEN SCIENCE CLOUD

Evaluation - Scoring

Summary:

	Score earned:	Fair level:
Findable:	6 of 7	moderate
Accessible:	2 of 3	moderate
Interoperable:	3 of 4	moderate
Reusable:	5 of 10	moderate

- Originally pass/fail and numerical score per metric
- Now maturity levels for each metric and principle
- Final FAIRness result:
  - Overall FAIR maturity level
  - total score (% of max)

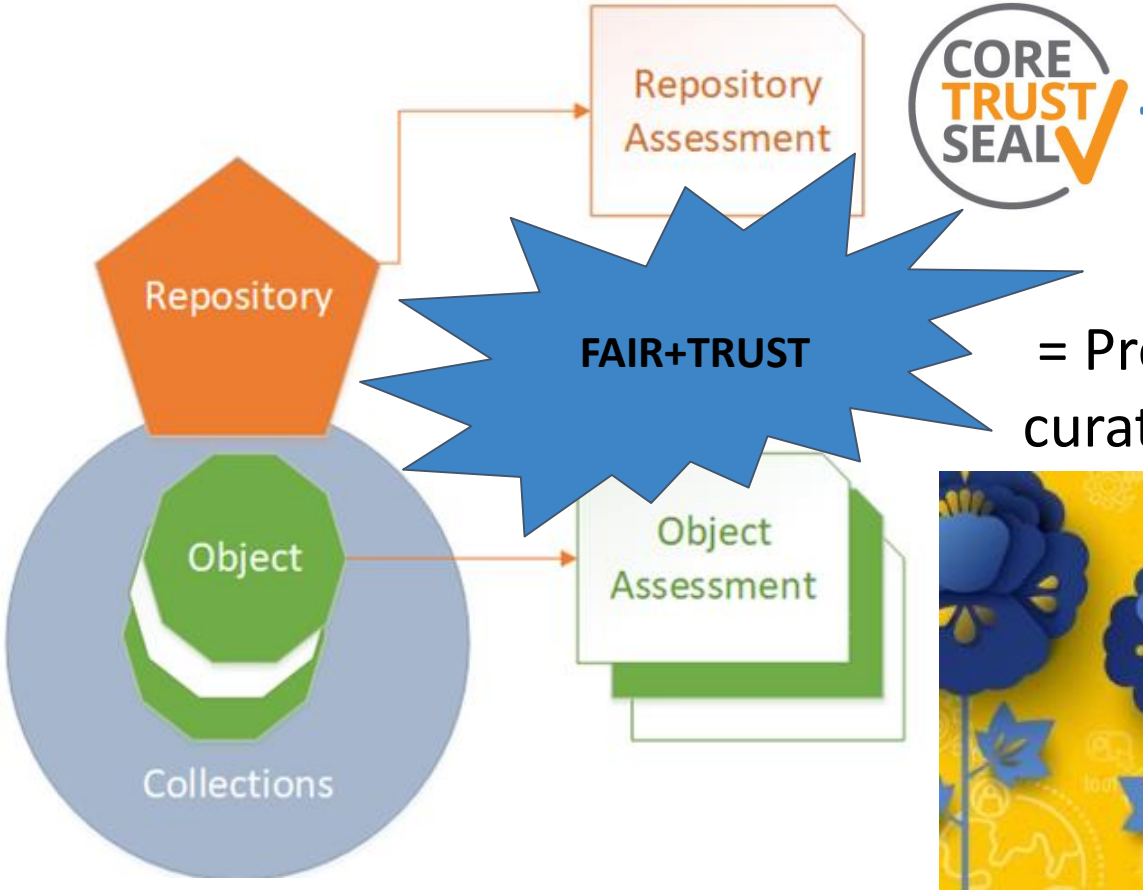
CTS FAIR CMM

- defined
- managed
- initial

F-UJI Maturity Levels

- advanced
- moderate
- initial
- incomplete

Unifying repository and object assessment



CORE TRUST SEAL +FAIR

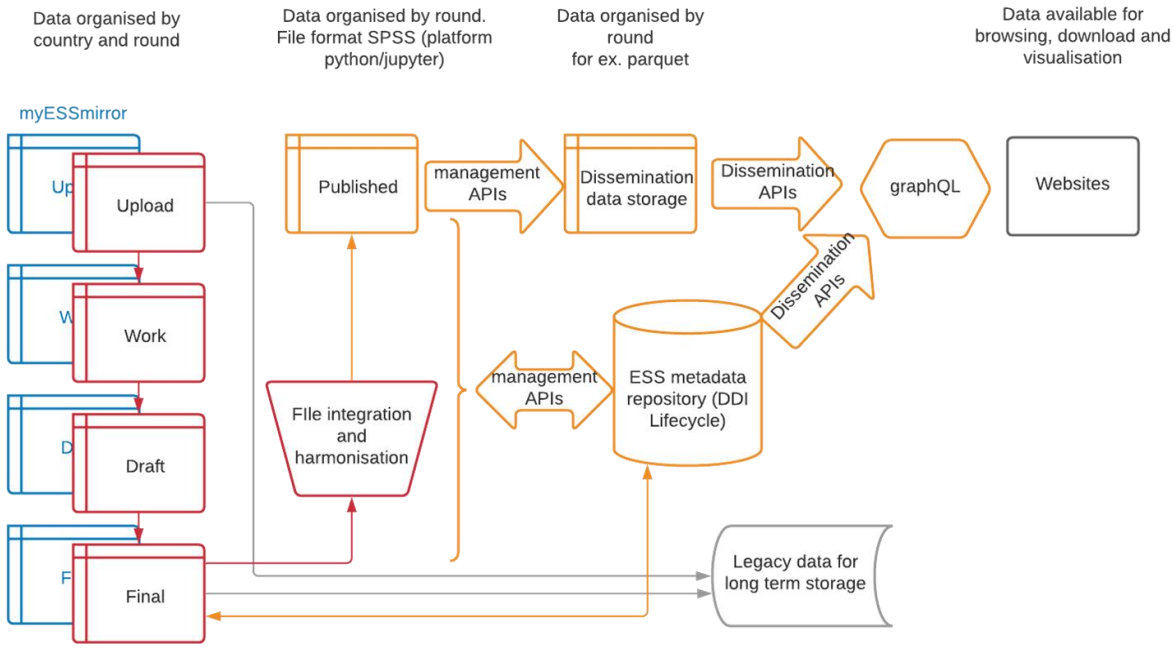
10 data repositories

= Preservation, curation, data

EUROPEAN OPEN SCIENCE CLOUD

Making cross-national survey data FAIR in ESS

- Integrate data publishing with DataCite DOI
- Make bespoke advanced landing pages with rich functionalities
- Landing pages are interoperate with the other infrastructure elements such as the repositories and the APIs
- Enables the content to be accessed in a FAIR way



European Social Survey

SSHOC

EUROPEAN OPEN SCIENCE CLOUD

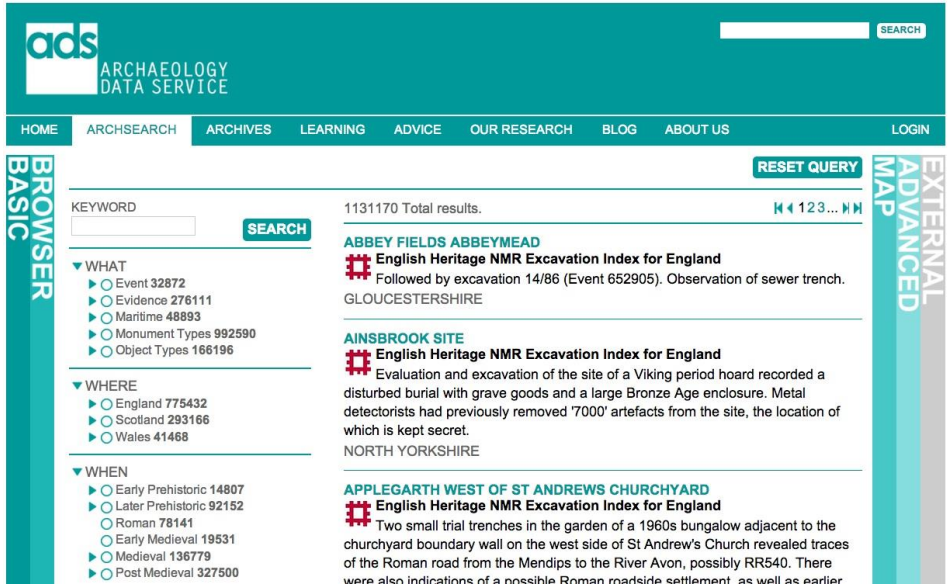
Automated FAIR assessment using F-UJI tool

F-UJI | Automated FAIR Data Assessment Tool

F-UJI is a service based on REST, piloting a programmatic assessment of the FAIRness of research datasets

FAIRSFAIR
Fostering Fair Data Practices in Europe

Apply F-UJI to the ADS ArchSearch catalogue which indexes over 1.3 million metadata records



SSHOC

EUROPEAN OPEN SCIENCE CLOUD

Example results

A1.1 The protocol is open, free, and universally implementable

External Qualitative Assessment

- The ADS uses the HTTPS protocol
- The repository utilises open and free file-sharing services where files or datasets are too large for easy exchange using HTTPS.

Internal Qualitative Recommendation

- Recommendation A1.1: A clear policy of sharing large files and datasets using more open services.

F-UJI Automated Assessment

Result	Comments	Next Step
Score: 1.0-1.0 of 1	OK	

I2. (Meta)data use vocabularies that follow FAIR principles

External Qualitative Assessment

- The ADS uses a variety of sustainable, open vocabularies to qualitatively classify and identify resources and datasets

Internal Qualitative Recommendation

- Investigate the FAIRness of vocabularies used by ADS
- Consider a more wholesale / consistent implementation of thesauri at object level
- Request clearer documentation from depositors where data makes use of controlled vocabularies

F-UJI Automated Assessment

Result	Comments	Next Step
Score: 0.0-0.0 of 1	Whereas the service seems to use controlled vocabularies such as http://purl.org/heritagedata it seems not be used in the metadata detected by F-UJI.	Rec.: Use vocabularies in schema.org as discussed here: https://github.com/ESIPFed/science-on-schema.org/issues/27

Outcome

- Now working with ARIADNE infrastructure to see how inclusion of metadata enhances FAIRness
- Aggregating ADS data in ARIADNE should make it more FAIR and interoperable as it will be automatically mapped to CIDOC CRM
- Also creating a workflow to address FAIR data quality – recommendations from collection development staff and technical team

EOSC Nordic FAIR developments

Tested F-UJI tool on 100 repositories

- 24% of the sample lacked PIDs
- 30% had no support for machine-actionable metadata
- Need for more controlled vocabs → M4M workshops
- Noticeable higher FAIR scoring among repositories being run on established platforms (Dataverse, Figshare, and others)
- Even higher scoring among CoreTrustSeal certified repositories

https://doi.org/10.5281/zenodo.5226082

Development of Service Interoperability Framework for Nordic context

https://eosc-nordic.atlassian.net/wiki/spaces/EN/pages/1276149763/EOSC-Nordic+Service+Interoperability+Framework

EOSC NORDIC

EUROPEAN OPEN SCIENCE CLOUD

Do I-PASS for FAIR

- Self-assessment tool for research organsiations to see how well they support researchers to be FAIR
- Covers policy, services, skills, incentives & adoption
- Helps to develop a roadmap for supporting FAIR
- Identifies national or domain specific challenges

https://datascience.codata.org/article/10.5334/dsj-2021-030

Questions?

Thanks!

EUROPEAN OPEN SCIENCE CLOUD