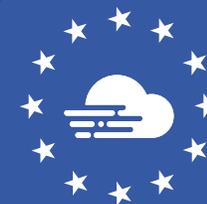


FAIR in the EOSC Association

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

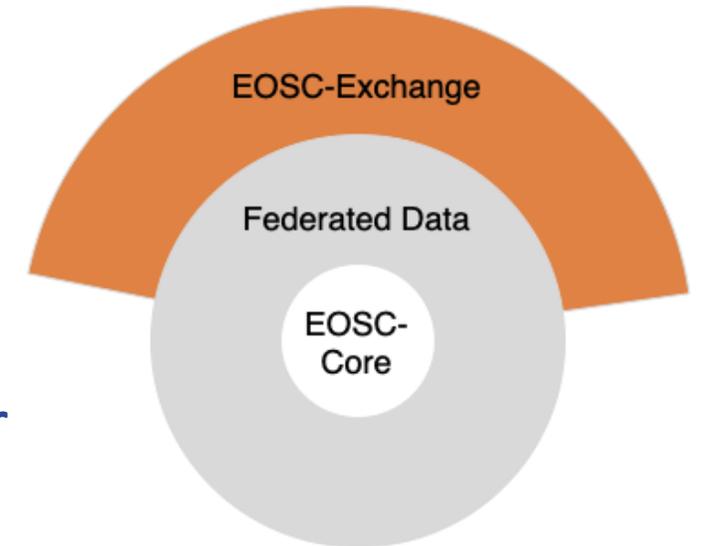
DeiC conference
3 November 2021



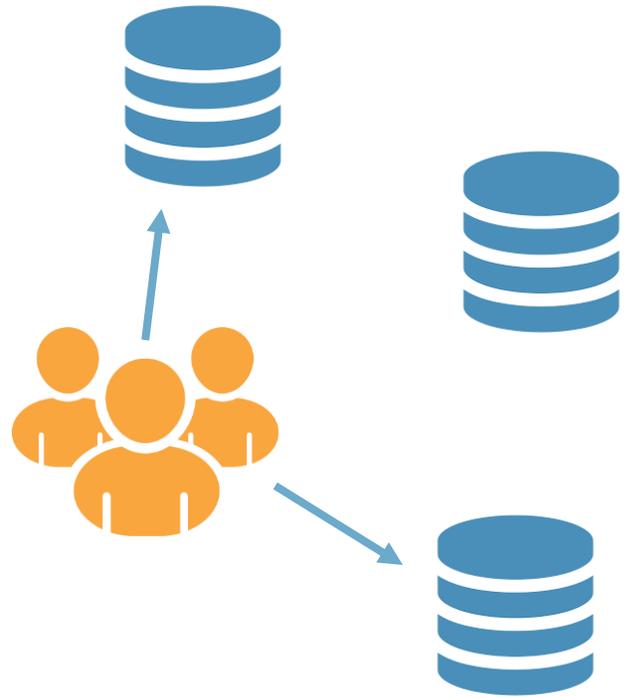
**EUROPEAN OPEN
SCIENCE CLOUD**

What is EOSC?

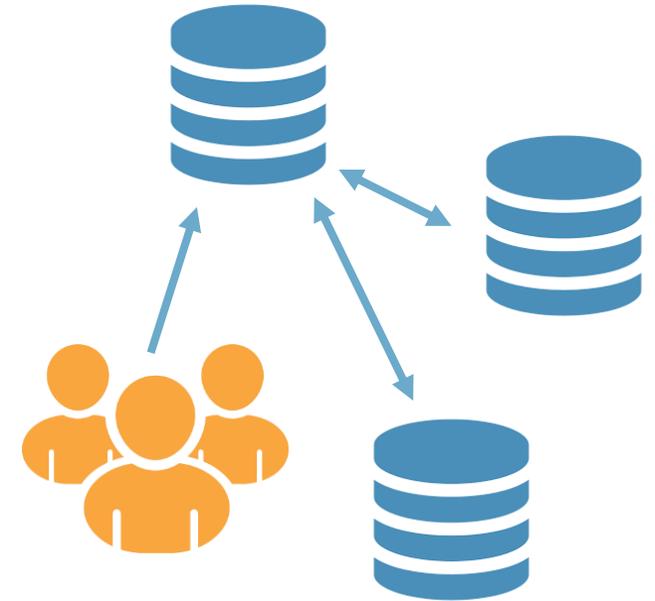
- ★ A web of FAIR data and services
- ★ Federation of eInfra and research infrastructures
- ★ Environment in which data can be brought together with services to perform analyses and address societal challenges



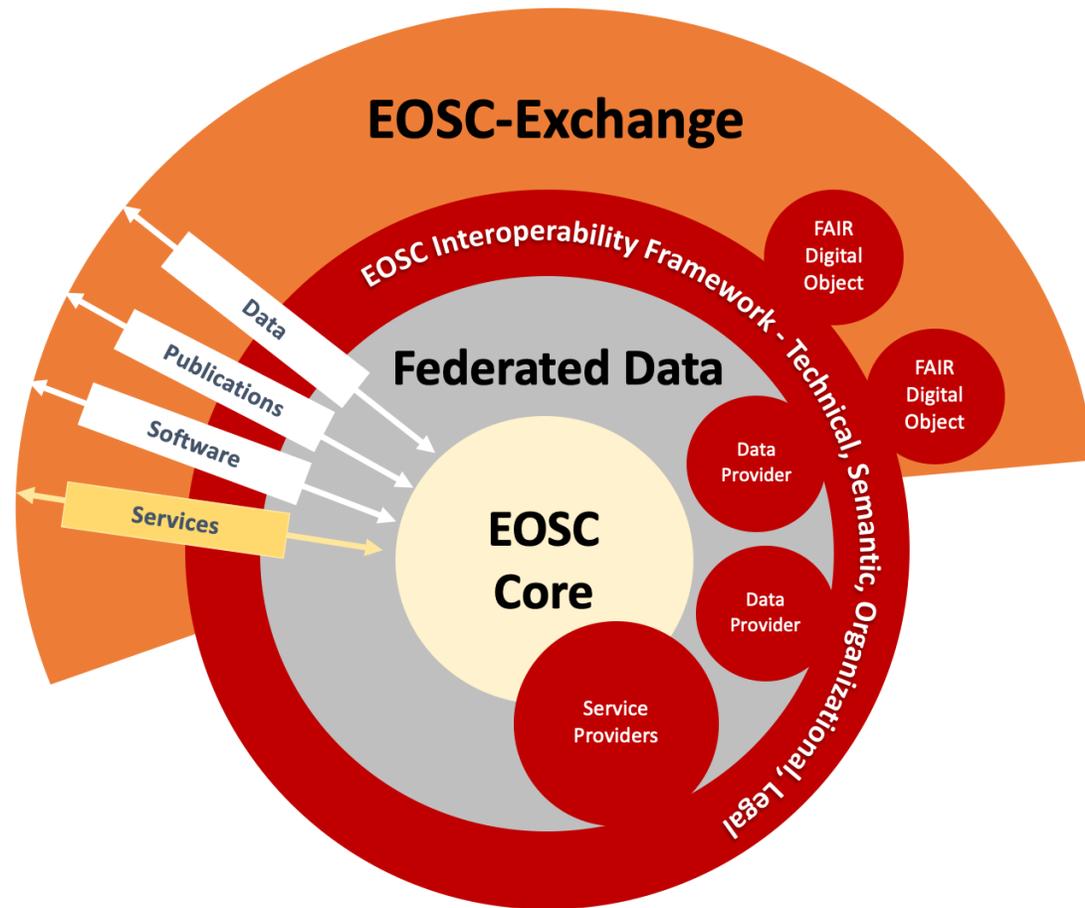
Aims to enable multidisciplinary discovery & use



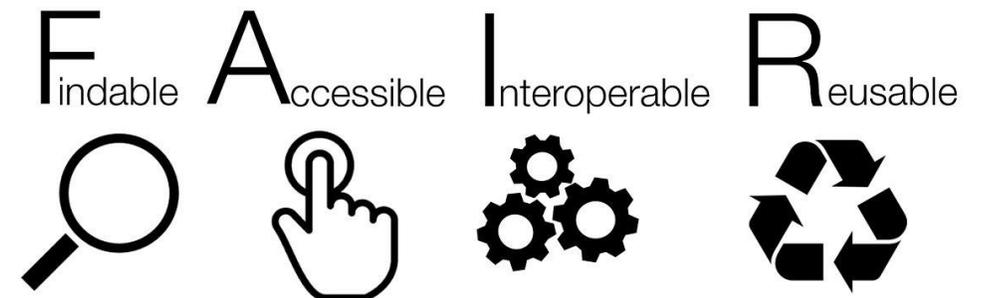
Disconnected silos to a federated infrastructure providing added value to researchers



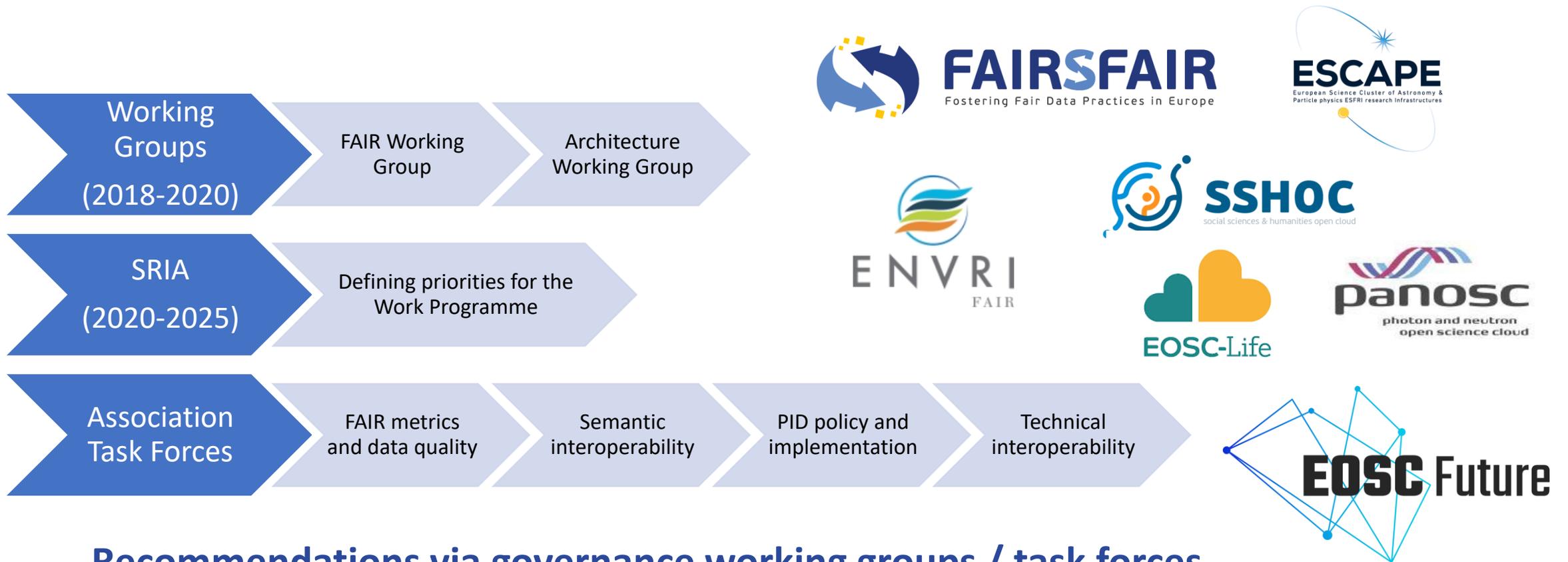
FAIR is central to principles in EOSC



- 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



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>

Task Force Leads

- Oscar Corcho
- Krzys Kurowski

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



FAIR practice

PID policy



Task Force Leads

- Peter Wittenberg
- Rachael Kotarski

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

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>



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

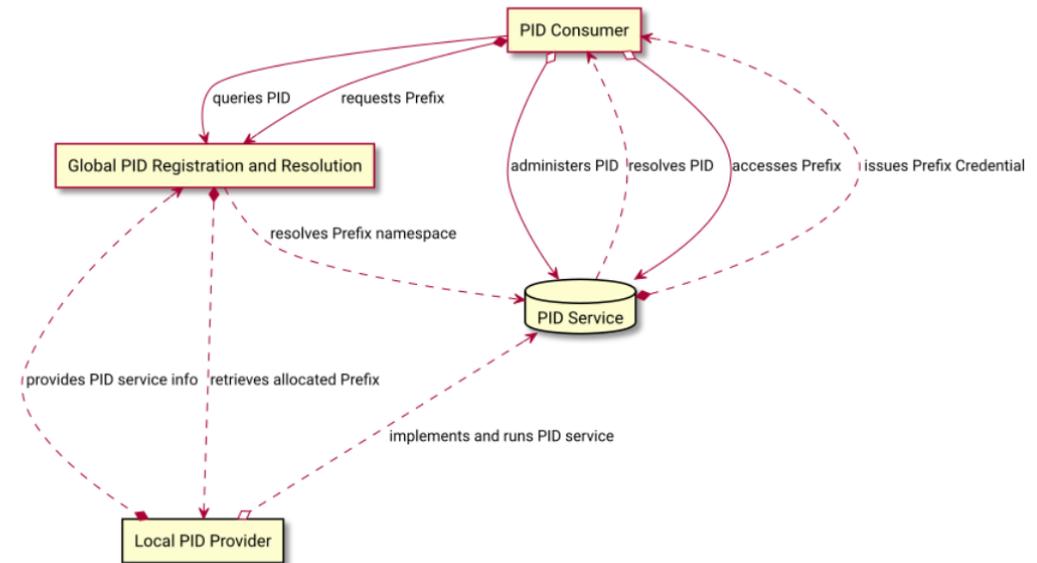
“The EOSC PID Policy identifies that an **ecosystem of PID infrastructures is needed** to support a wide variety of scientific applications and offer sufficient flexibility and capacity. The ARDC PID portfolio also aims to cover off important elements of the research enterprise (**people, organisations, 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

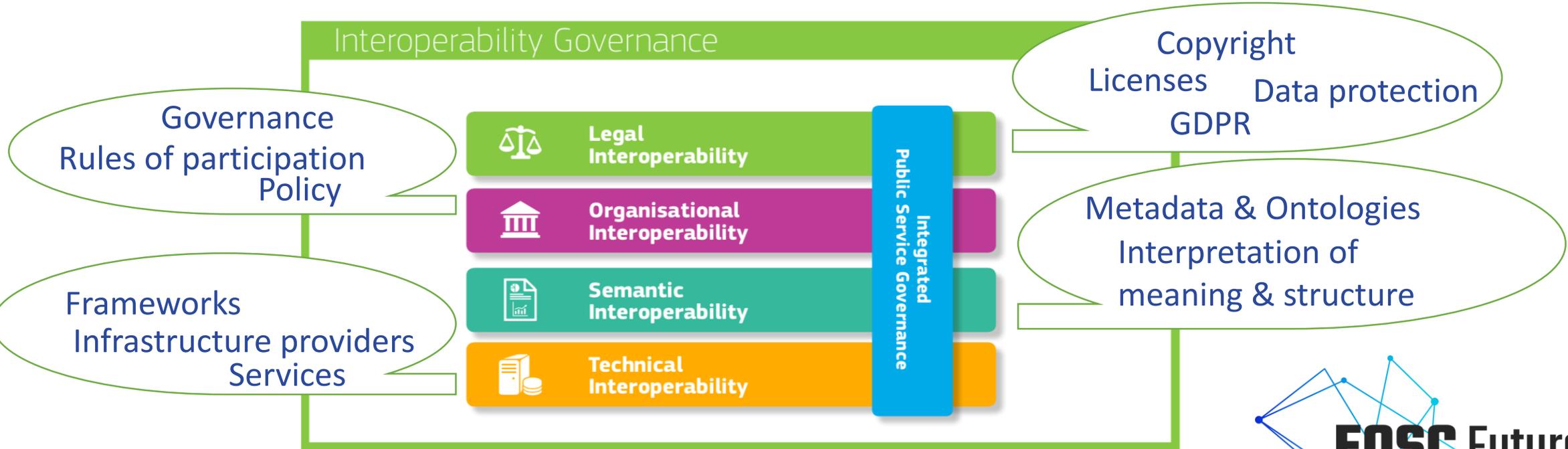
Three highest ranked priorities:

- ★ 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>

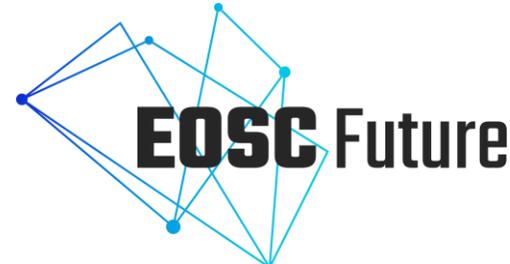


EOSC Interoperability Framework



The European Interoperability Framework four levels of interoperability

<https://doi.org/10.2777/620649>



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	<ul style="list-style-type: none">• Metadata is provided to allow discovery• Metadata includes the identifier for the data	<ul style="list-style-type: none">• Metadata is offered in such a way that it can be harvested and indexed	<ul style="list-style-type: none">• Metadata is guaranteed to remain available, after data is no longer available
Licence	<ul style="list-style-type: none">• Metadata includes information about the license under which the data can be reused	<ul style="list-style-type: none">• Metadata refers to a standard reuse license	<ul style="list-style-type: none">• Metadata refers to a machine-understandable reuse licence
Standards		<ul style="list-style-type: none">• Data/Metadata complies with a community standard	<ul style="list-style-type: none">• 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

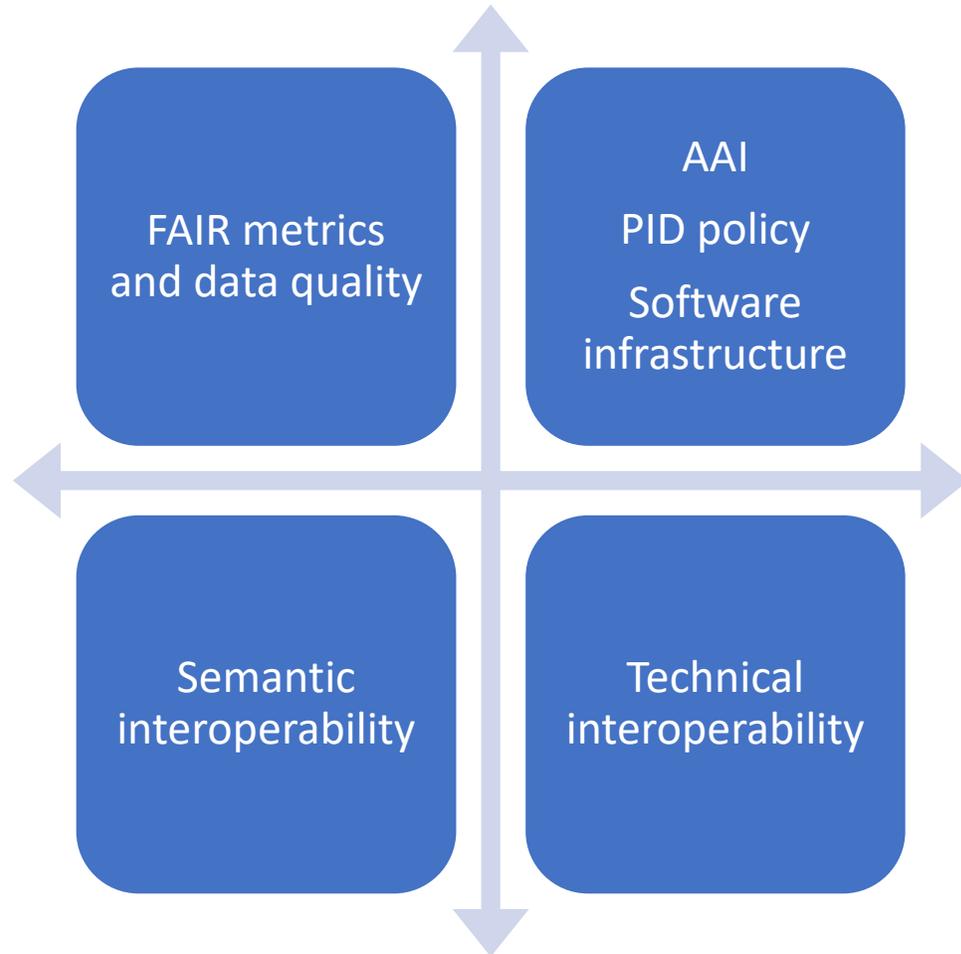
HORIZON-INFRA-2021-EOSC-01-02

- ★ 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



Characteristics of Environmental Infrastructures

Societal challenges need multidisciplinary methods

Specialized observation and analysis platforms

High importance to society, economy and resilience

Life

Air

Answering societal needs is only possible via collaboration

Observations often unique - huge datasets

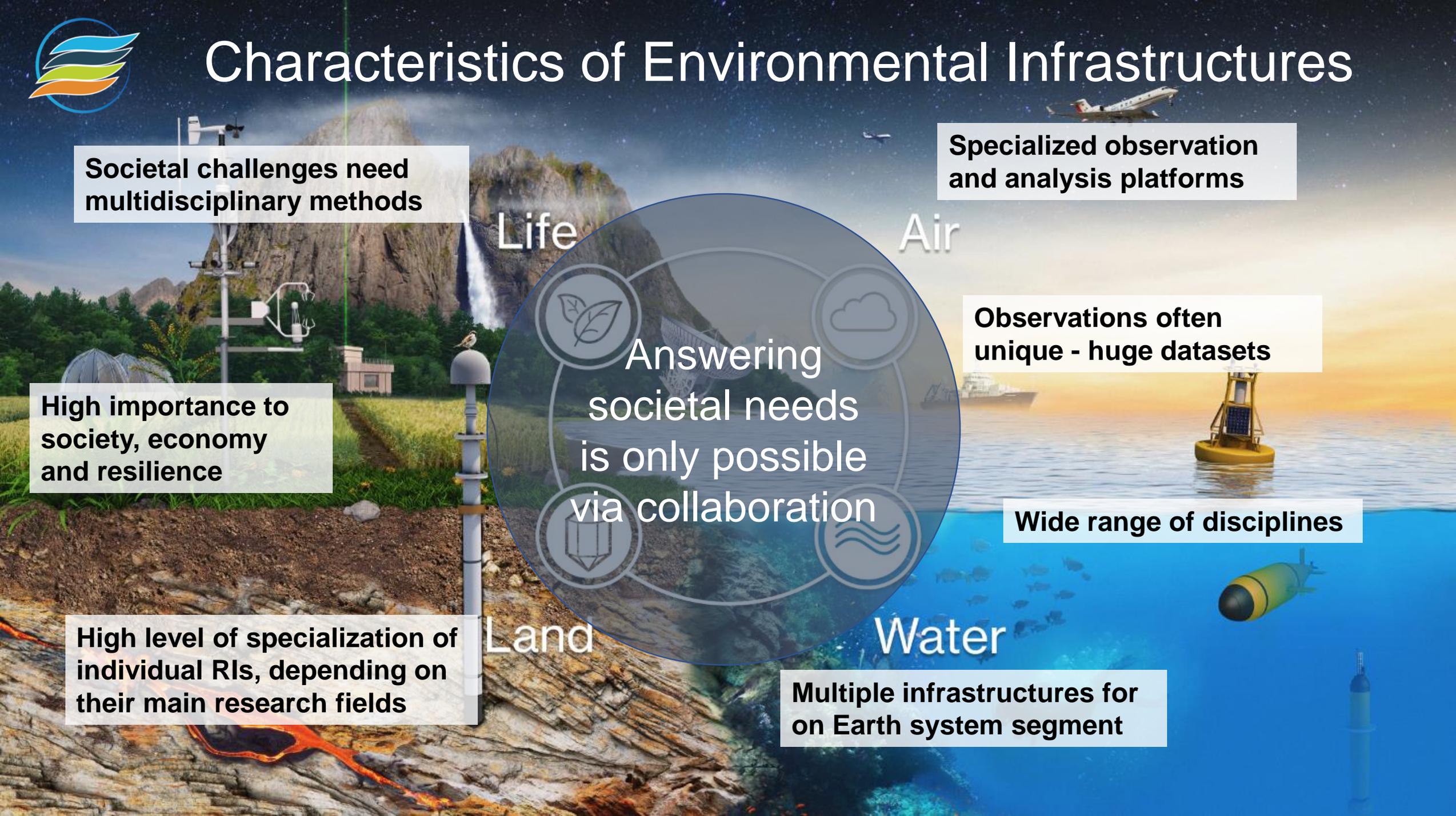
Wide range of disciplines

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

Land

Water

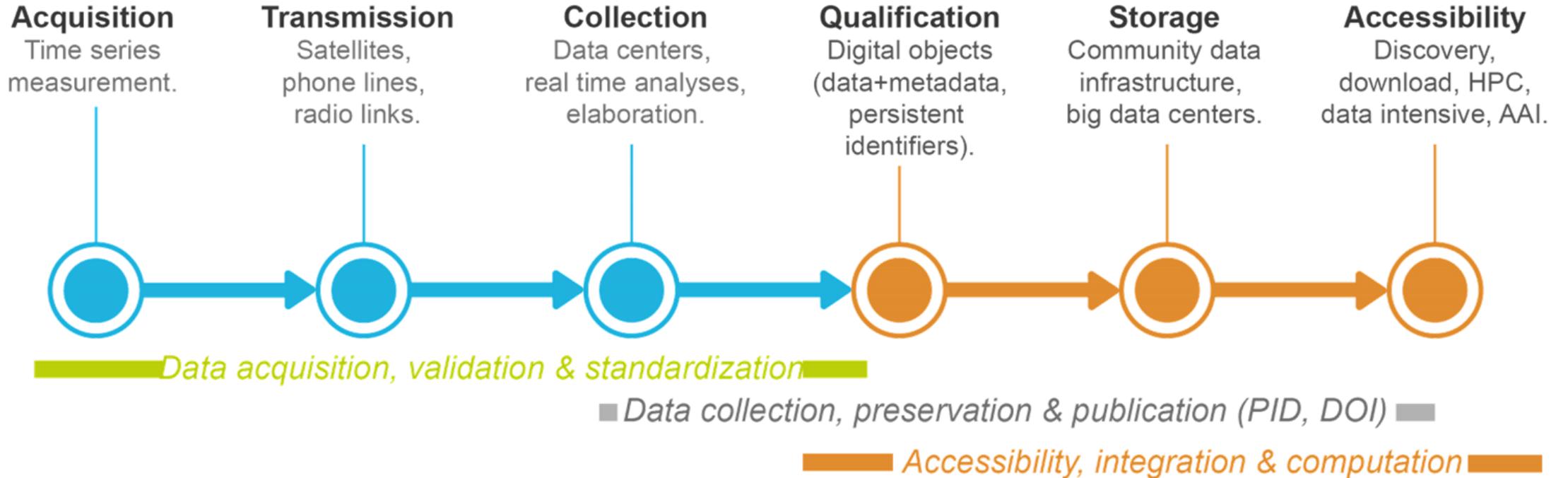
Multiple infrastructures for on Earth system segment





Data Lifecycle Challenges in the ENVRI Domain

Services



Standard
Operating
Procedures

MetaData formats
Naming conventions
Data analysis
workflows

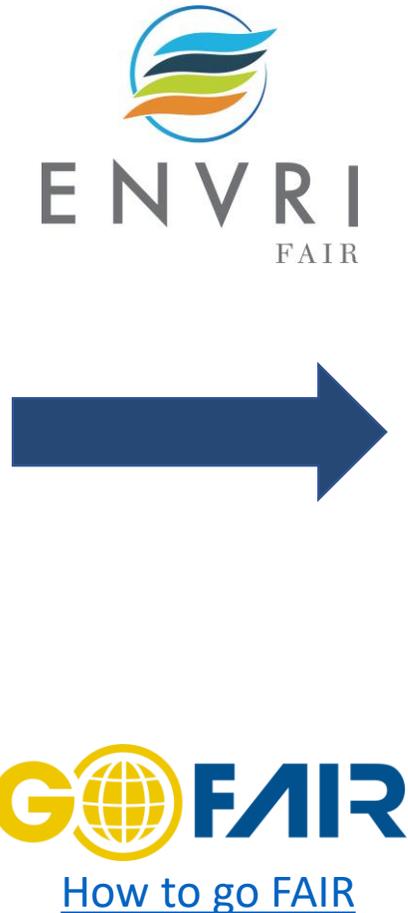
Use of PIDs
Workflow
management
Provenance

Trusted
repositories
Sustainable
infrastructures

Vocabularies
Machine access
AAI
Licenses

FAIR Assessment Methods used in ENVRI-FAIR

FAIR Assessment Questionnaire



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





ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

Examples of implementing FAIR from ESCAPE

Mark Allen for the CEVO Work Package (WP4)

CNRS-ObAS Observatoire astronomique de Strasbourg

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

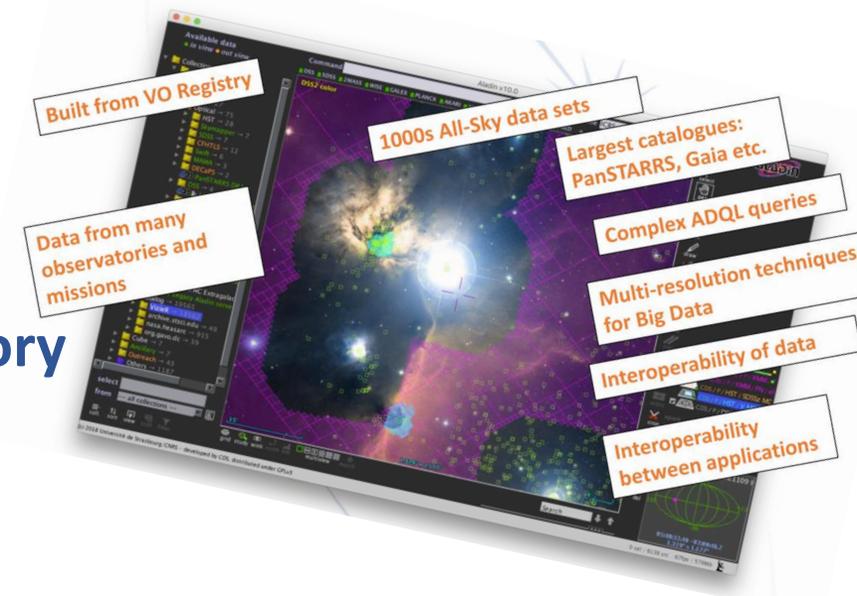
See Astronomy use case in SRIA, and EOSC Interoperability Framework

EOSC to enable next steps of the astronomical Virtual Observatory

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



IVOA

22,234 datasets found for "IVOA"

Order k

ESO TAP_OBS: a TAP service to browse and access raw and red
TAP_OBS is the ESO Science Archive TAP endpoint for observations (raw and red) against our database tables, inspect various metadata, and upload your own data. It is thus the VO's premier way to access public data holdings.

UCL DaCHS server TAP service
The UCL DaCHS server's TAP endpoint. The Table Access Protocol (TAP) lets you execute queries against our database tables, inspect various metadata, and upload your own data. It is thus the VO's premier way to access public data holdings.

Dataset

Communities

The VO @ ASTRON TAP service

The The VO @ ASTRON's TAP endpoint. The Table Access Protocol (TAP) lets you execute queries against our database tables, inspect various metadata, and upload your own data. It is thus the VO's premier way to access public data holdings.

Tables exposed through this endpoint include: main from the lbcas schema, main, mom0 from the sauron schema, img_main, main from the mvf schema, columns, groups, key_columns, keys, schema, main, msssvf_img_main from the tap_schema schema, hetdex_images, img_main from the hetdex schemas, tables from the tap_schema schema, obscure from the ivoa schema.

ADQL

Catalogs

Virtual observatory

Identifier

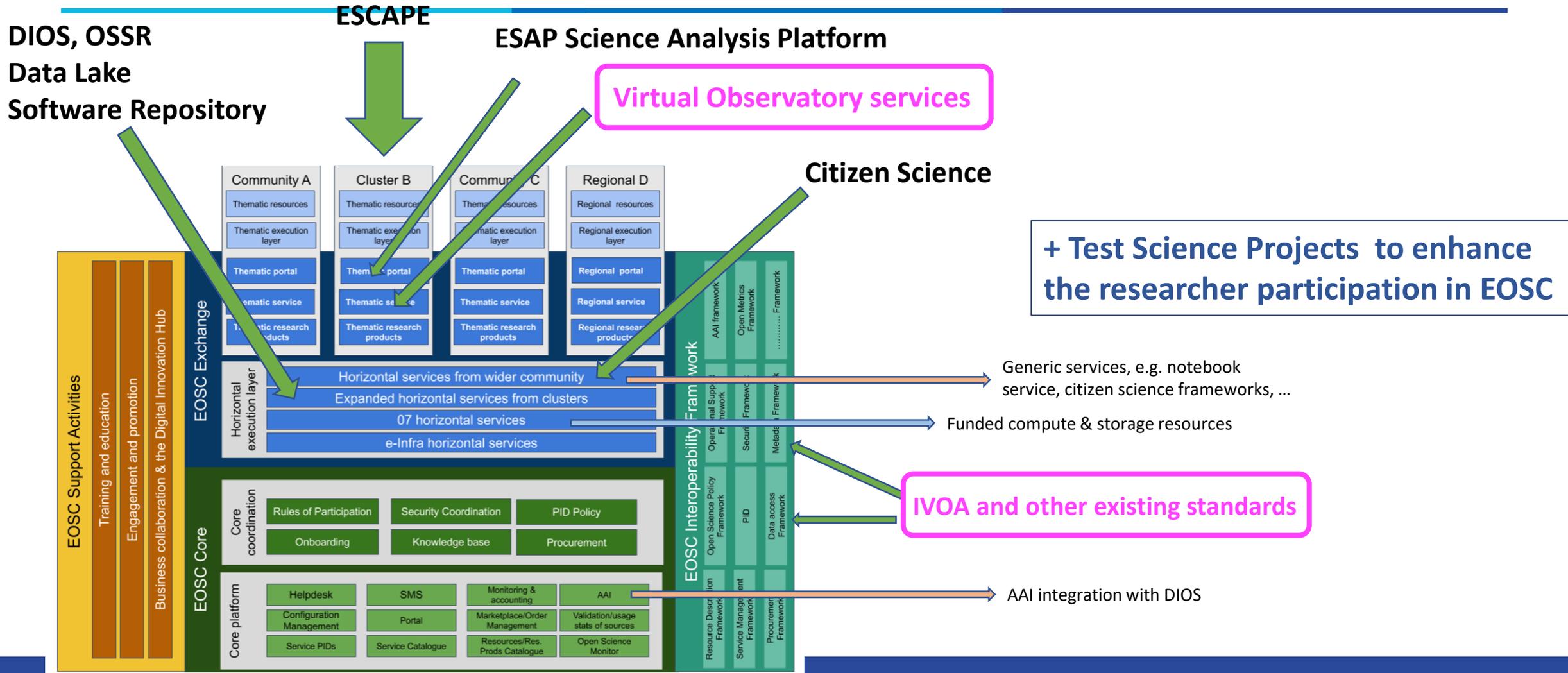
Source

Metadata Access

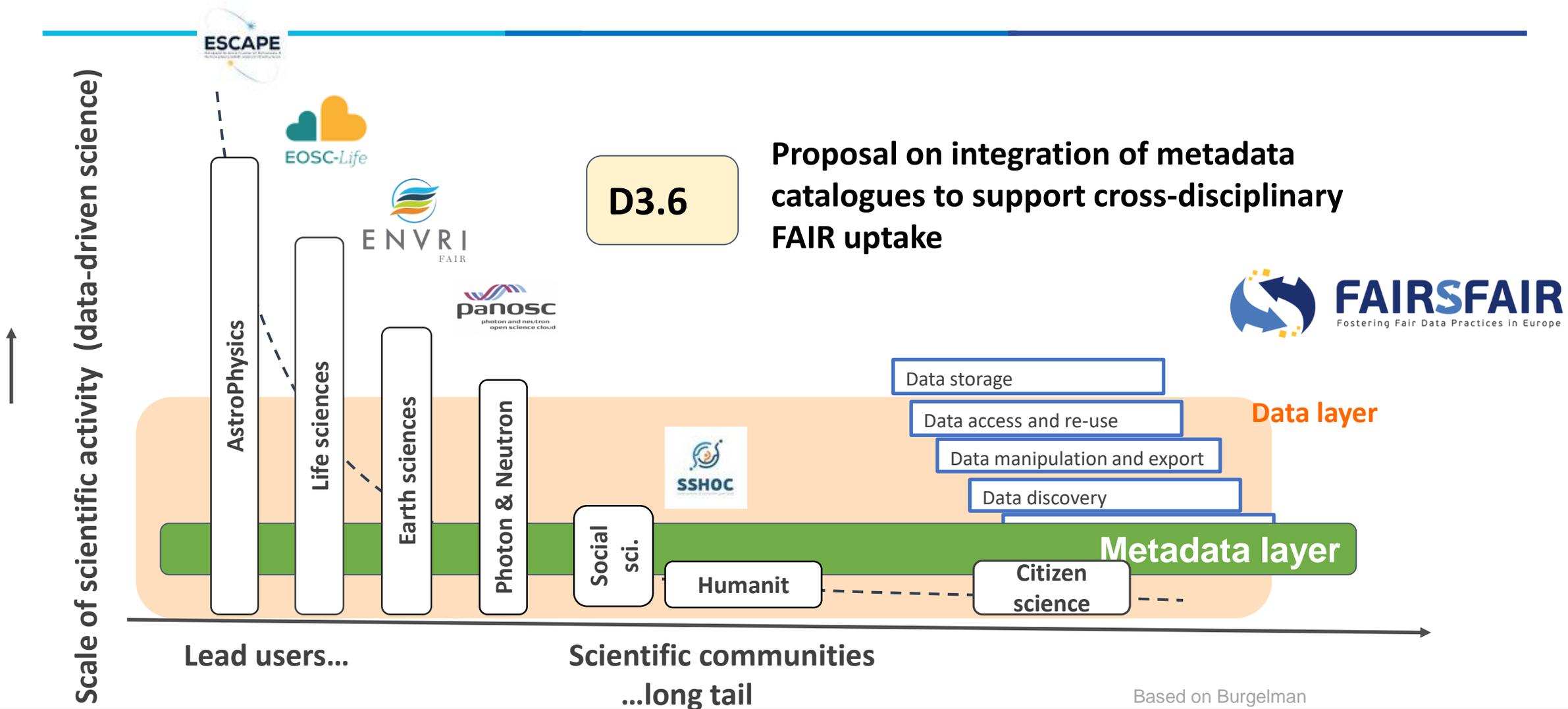
https://vo.astron.nl/_system_/tap/run/info
<http://dc.g-vo.org/rr/q/pmb/verb=GetP>

Provenance

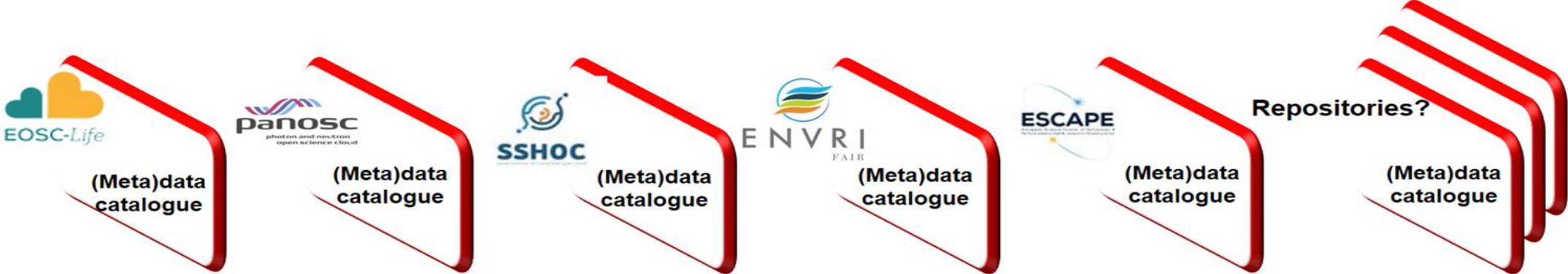
ESCAPE connecting to EOSC... a work in progress



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

The screenshot displays the F-UJI web interface. At the top, there is a search bar containing the path `/fuji/api/v1/openapi.json` and an **Explore** button. Below this, the page title is **F-UJI 1.0.0 OAS3**. The main content area includes a description: "A Service for Evaluating Research Data Objects Based on FAIRsFAIR Metrics." and a note about funding: "This work was supported by the FAIRsFAIR project (H2020-INFRAEOSC-2018-2020 Grant Agreement 831558)." There are links for "Contact the developer", "MIT License", and "Find out more about F-UJI".

On the left side, there is a "Servers" section with a dropdown menu showing `/fuji/api/v1`. Below this is a "FAIR object" section with the text "FAIRness assessment of a data object".

The main part of the interface shows a list of API endpoints. The first one is a **POST** endpoint `/evaluate`. A red arrow points from this endpoint to a detailed view of the response. The second endpoint is a **GET** endpoint `/metrics` with the description "Return all metrics and their definitions".

The detailed view of the `/evaluate` endpoint shows a **200** status code and a **Response body** containing JSON data. The JSON data includes:

```
{
  "metric_identifier": "FsF-F1-02D",
  "metric_name": "Persistent identifier",
  "output": {
    "pid": "https://doi.org/10.1594/PANGAEA.902845",
    "pid_scheme": "doi",
    "resolvable_status": true,
    "resolved_url1": "https://doi.pangaea.de/10.1594/PANGAEA.902845"
  },
  "passed": true,
  "score": {
    "earned": 1,
    "total": 1
  },
  "test_debug": [
    "INFO: Persistence identifier scheme - doi",
    "INFO: Retrieving page http://doi.org/10.1594/PANGAEA.902845",
    "INFO: Request status code - 200",
    "INFO: Found HTML page"
  ]
},
{
  "id": 3,
  "metric_identifier": "FsF-F2-01M",
  "metric_name": "Descriptive (core) metadata",
  "output": {
    "core_metadata_found": {
      "creator": [
        "Bärfuss, Konrad",

```

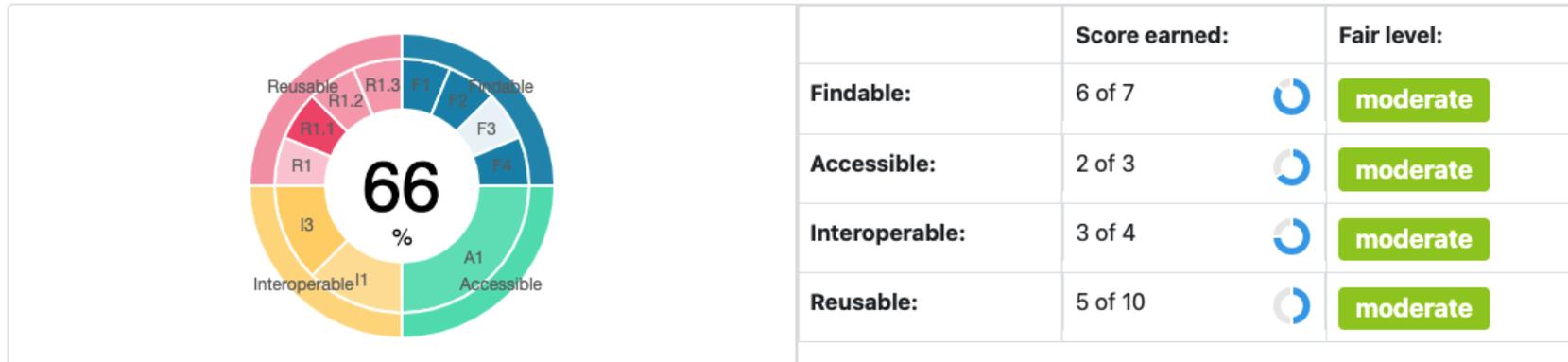
Below the response body, there are **Response headers** listed:

```
content-length: 5116
content-type: application/json
date: Fri, 24 Apr 2020 17:14:06 GMT
server: Werkzeug/1.0.0 Python/3.7.6
```

There is an **Authorize** button next to the endpoint list and a **Download** button next to the response body.

Evaluation - Scoring

Summary:



- 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)

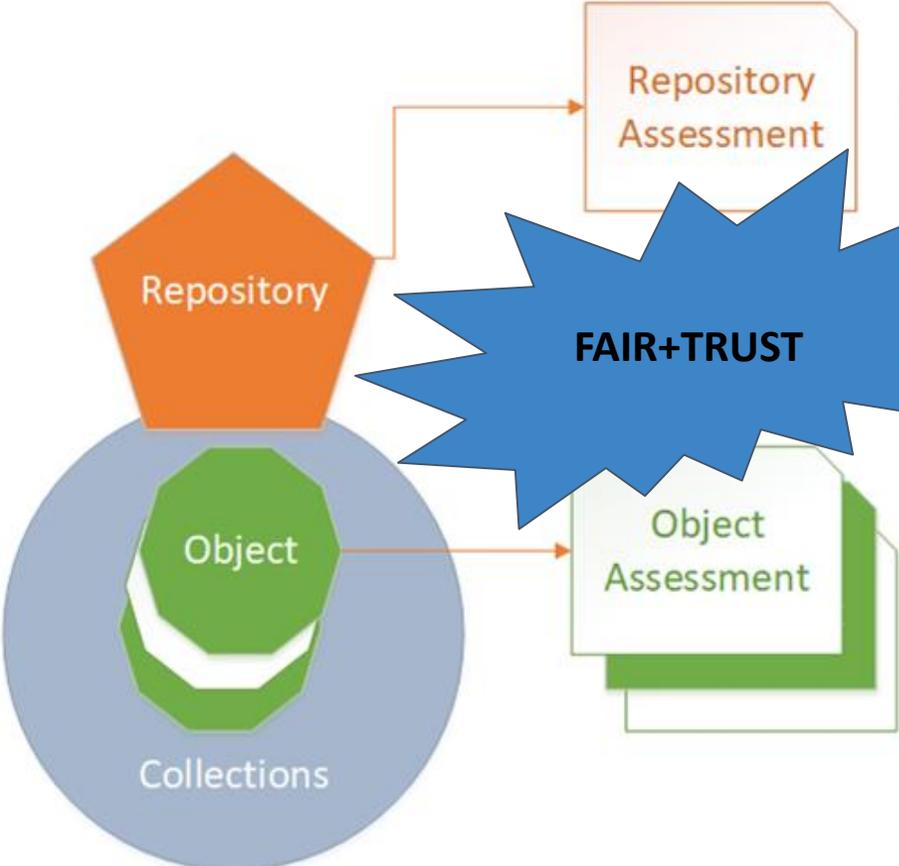
CTS FAIR CMM



F-UJI Maturity Levels



Unifying repository and object assessment



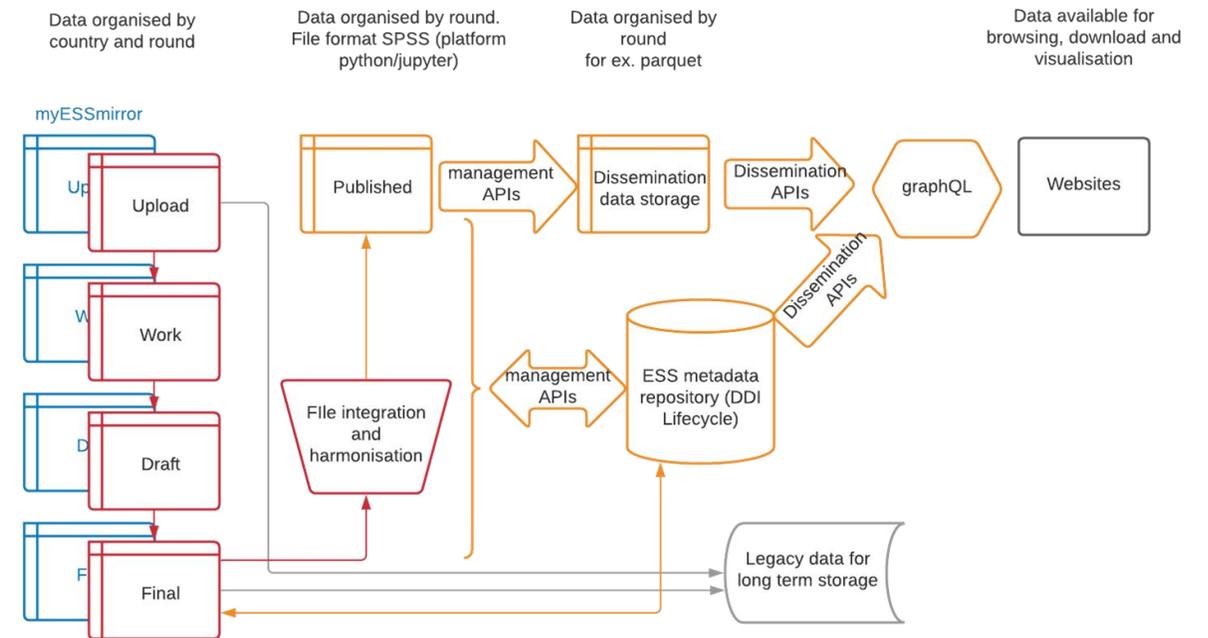
10 data repositories

= Preservation, curation, data



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 interoperable with the other infrastructure elements such as the repositories and the APIs
- Enables the content to be accessed in a FAIR way



Automated FAIR assessment using F-UJI tool



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



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

ads ARCHAEOLOGY DATA SERVICE

HOME ARCHSEARCH ARCHIVES LEARNING ADVICE OUR RESEARCH BLOG ABOUT US LOGIN

RESET QUERY

KEYWORD SEARCH

1131170 Total results. << 123 >>

ABBEY FIELDS ABBEYMEAD
English Heritage NMR Excavation Index for England
Followed by excavation 14/86 (Event 652905). Observation of sewer trench. GLOUCESTERSHIRE

AINSBROOK SITE
English Heritage NMR Excavation Index for England
Evaluation and excavation of the site of a Viking period hoard recorded a disturbed burial with grave goods and a large Bronze Age enclosure. Metal detectorists had previously removed '7000' artefacts from the site, the location of which is kept secret. NORTH YORKSHIRE

APPLEGARTH WEST OF ST ANDREWS CHURCHYARD
English Heritage NMR Excavation Index for England
Two small trial trenches in the garden of a 1960s bungalow adjacent to the churchyard boundary wall on the west side of St Andrew's Church revealed traces of the Roman road from the Mendips to the River Avon, possibly RR540. There were also indications of a possible Roman roadside settlement, as well as earlier

BROWSER BASIC ADVANCED MAP EXTERNAL



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>



Do I-PASS for FAIR

- ★ Self-assessment tool for research organisations 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>

The screenshot shows the 'DO I-PASS FOR FAIR?' self-assessment tool interface. At the top right, there is a green box with the text 'Self assessment tool to measure the FAIR-ness of an organization'. Below this, there are three buttons: 'BEGINNER' (highlighted in green), 'INTERMEDIATE', and 'ADVANCED'. The main content area is dark blue and contains the heading 'DOES YOUR ORGANIZATION...' followed by five numbered questions:

- 1 POLICY**
...have a FAIR research data policy?
- 2 SERVICES**
...have a DCC which provides services to allow research(ers) to comply with FAIR?
- 3 SKILLS**
...acknowledge that FAIR capacity building requires specific roles and skills?
- 4 INCENTIVES**
...have incentives for FAIR data?
- 5 ADOPTION**
...have adoption of FAIR?

At the bottom left, there is a green box with the text 'Find the checklist here DOI: 10.5281/zenodo.4080867'. At the bottom right, there is a logo for 'Landelijk Coördinatiepunt Research Data Management The data support collective'.

Questions?



Thanks!