SLICES infrastructure and services integration with EOSC and Open Science

Defining SLICES Interoperability Framework

Yuri Demchenko, Kishor Joshi

CCI, University of Amsterdam

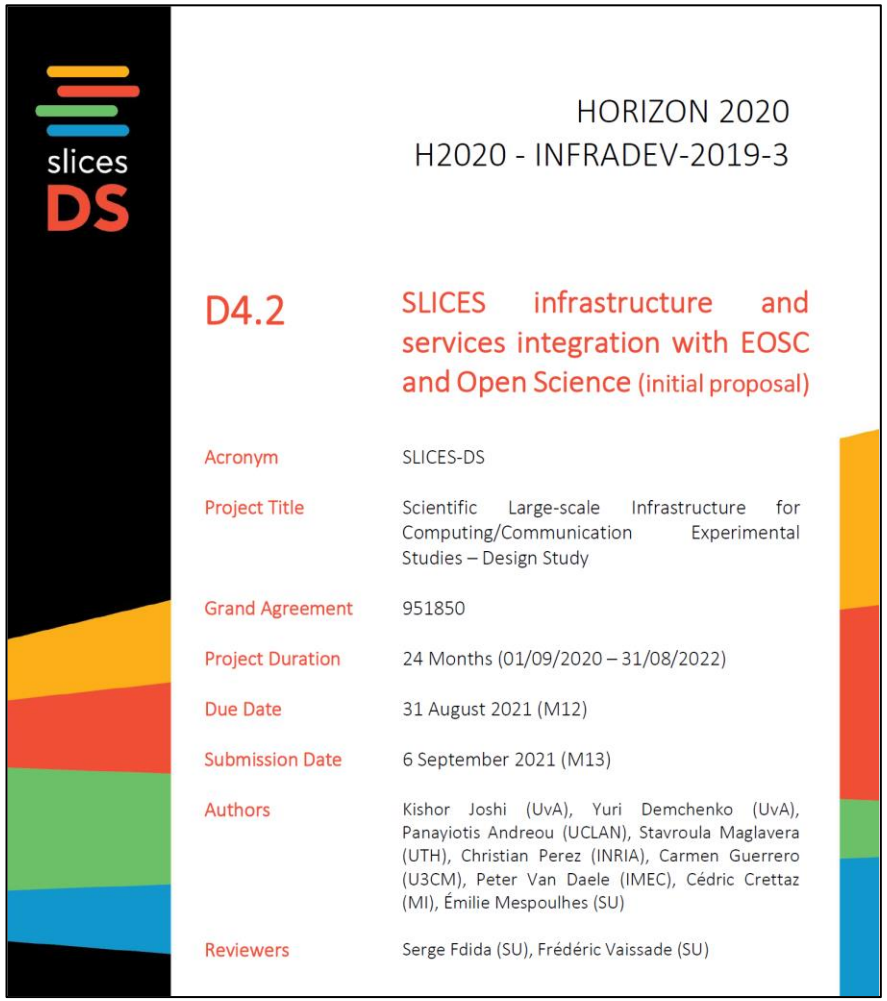
SLICES-RI Workshop

23 September 2021

Outline

- SLICES-DS WP4 Integration and compatibility with EOSC, FAIR and external RIs
  - Deliverable D4.2 SLICES infrastructure and services integration with EOSC and Open Science
  - Deliverable D4.3 Definition of the SLICES metadata profiles to support FAIR principles
- European Open Science Status Overview
  - Minimum Viable EOSC – core elements to enable EOSC and Federated Scientific Data Infrastructure
- EOSC Interoperability Framework
  - Components and technical view
- FAIR data principles: Technical aspects
- Discussion

Deliverable D4.2 SLICES infrastructure and services integration with EOSC and Open Science – *SLICES Interoperability Framework*


slices DS

HORIZON 2020
H2020 - INFRADEV-2019-3

D4.2

SLICES infrastructure and services integration with EOSC and Open Science (initial proposal)

Acronym	SLICES-DS
Project Title	Scientific Large-scale Infrastructure for Computing/Communication Experimental Studies – Design Study
Grand Agreement	951850
Project Duration	24 Months (01/09/2020 – 31/08/2022)
Due Date	31 August 2021 (M12)
Submission Date	6 September 2021 (M13)
Authors	Kishor Joshi (UvA), Yuri Demchenko (UvA), Panayiotis Andreou (UCLAN), Stavroula Maglavera (UTH), Christian Perez (INRIA), Carmen Guerrero (U3CM), Peter Van Daele (IMEC), Cédric Crettaz (MI), Émilie Mespoulhes (SU)
Reviewers	Serge Fdida (SU), Frédéric Vaissade (SU)

slicesRI

Deliverable D4.2 SLICES infrastructure and services integration with EOSC and Open Science – *SLICES Interoperability Framework*

SLICES Interoperability Framework – Goals

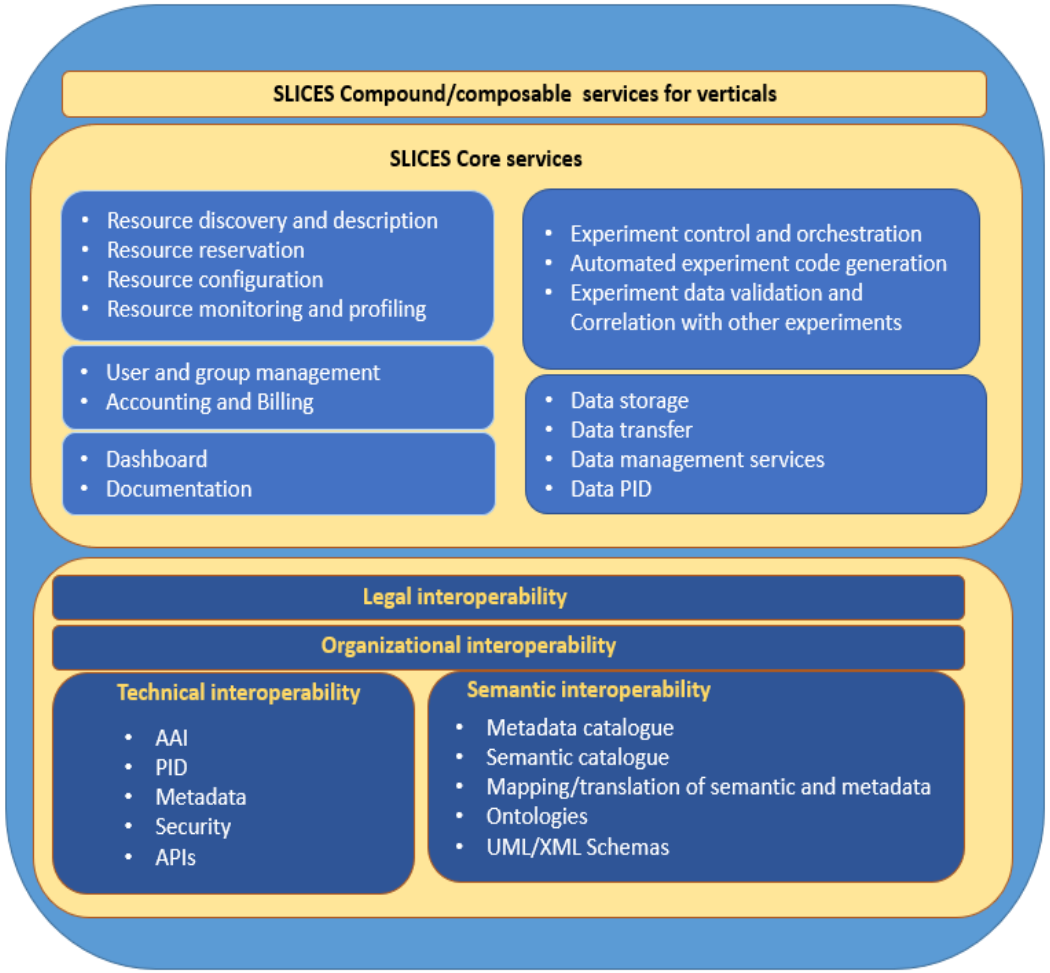
- *Requirements to SLICES interoperability and connection with EOSC*
- *What infrastructure services from EOSC can be re-used?*
- *How to access EOSC data resource and how to publish SLICES data to EOSC?*
- *What and how SLICES can contribute to EOSC?*

Why interoperate and connect with EOSC

- *EOSC initiatives are accepted by ESFRI and EC*
- *Mandatory in Horizon Europe proposals: DMP, FAIR, Open Science, PID/Zenodo,* ***ORE (new)*** *as well as liaison with EOSC*
- *EOSC is a model Federated Scientific Data Infrastructure and community*

slicesRI

Conceptual view of SLICES Interoperability Architecture



- Provides vision and roadmap to achieving interoperability with EOSC
- Some services can be used from EOSC, some services will require API with EOSC services of metadata mapping

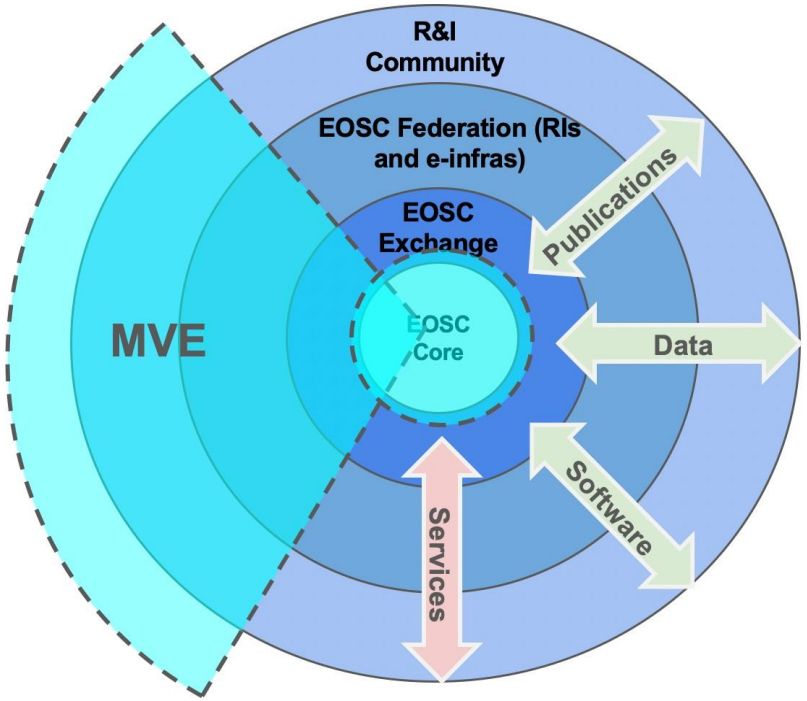
European Open Science Cloud (EOSC) – Overview

- EOSC is an overarching concept and framework to integrate existing RIs and facilitate information and data exchange between RI, organisations and researchers
  - First phase 2016-2020 with funded projects 2016-2022
    - 53 projects in total
- EOSC main projects and co-creation activities
  - EOSCpilot – Initial EOSC architecture and requirements (10 Mln, ended 2018)
  - EOSChub – Technical integration platform, RI marketplace and API/services directory (33 Mln, ended 31 March 2021)
  - EOSCsecretariat https://www.eoscsecretariat.eu/ (10 Mln, ends 31 October 2021)
    - Establishment of the Governance structure and EU EOSC association to be co-funded by EC and Member States (MS) Co-creation model and European Open Science Commons
  - SRIA (Strategic Research and Innovation Agenda) document (195 pages, Final version dated 15 Feb 2021) - https://www.eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf
- Built on experience of the past successful initiatives and project
  - EGEE and WLCG, EGI, RDA (co-founded by EC and NSF), GEANT/TERENA
- Can provides a model experience for future EU initiatives, such as GAIA-X European Federated Data Cloud

EOSC Conceptual Architecture

EOSC Docs

- EOSC-Core
  - Minimum architecture elements to enable the Federation
- EOSC-Exchange
  - Evolving Federation to serve the needs of research communities
  - Widening to the general public and the private sector
- EOSC Federation (RIs and e-Infra)
- Research and Innovation Community
- Minimal Viable EOSC (aka MVE)
  - Minimum Federation to bring value to users



EOSC Status: Minimum Viable EOSC and EOSC Core

- Architecture defining (infrastructure) components
  - Metadata Framework (FAIR data enabling services)
  - PID framework/Infrastructure and service
  - Federated Authentication, Authorisation Interoperability (AAI) Framework
  - Data Access framework
  - Service Management and Access framework
  - Open Metrics Framework

- Policy and Governance
  - Shared Open Science policy framework
  - A minimum legal metadata framework as part of the FAIR compliance framework
  - An open metrics framework

- Portal providing web access to the EOSC services
- EOSC is an important stage in the European RI integration

Demand for modern RI platform using recent development by industry (for future technology exchange)

EOSC challenges: to incorporate recent technology development into ERA

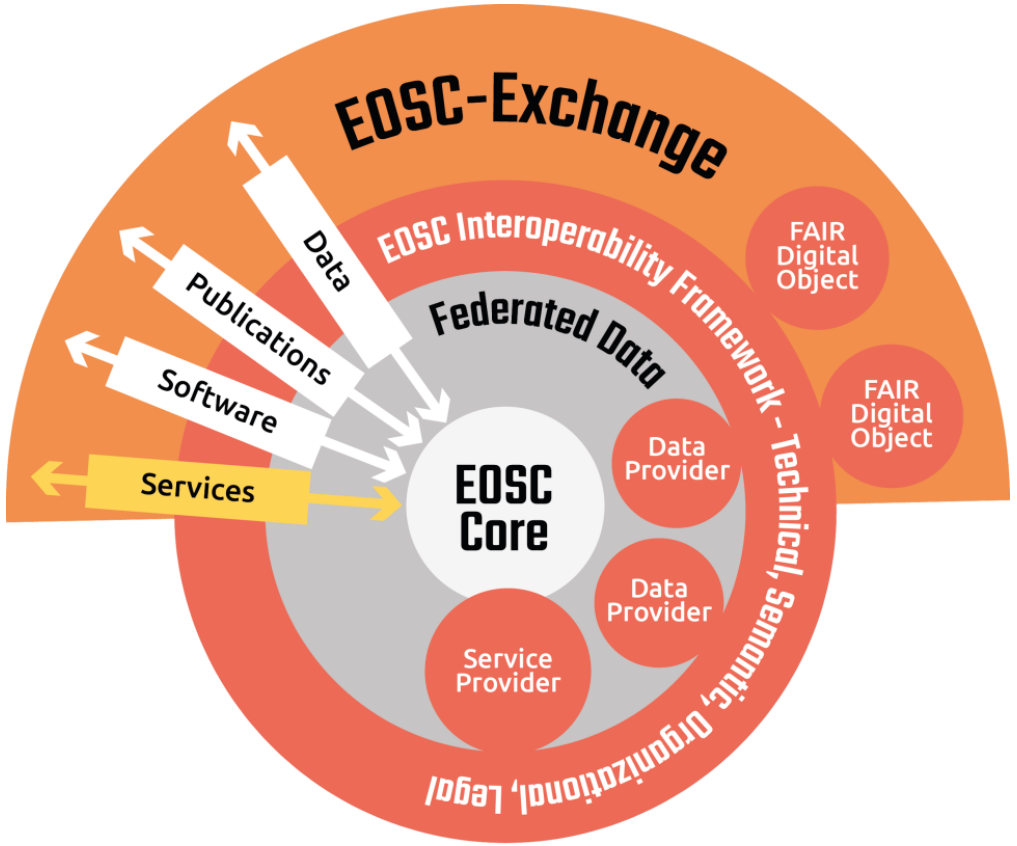
Future EOSC Development and SLICES: Timeline e-RI evolution and SLICES positioning

SLICES

RI Type (evolution stage)	Centralised 1994-1996	Interconnected 2004-2006	Distributed 2011-2012	Federated 2016-2018	EOSC-1 2020-2022	EOSC-2 (future) 2025+
Definition	Institutions based, centralised facility	Multi-institutions, interconnected	Large distributed facilities, domain or experiment oriented	Federated RIs supporting inter-domain cooperation and data exchange	Interoperable (European) RI, FAIR RI	Virtualised Pan-European RI platform as a Service and ecosystem (PRIaaS)
Network & Compute	Mainframe, variety of protocols, Advent of Internet, web, email	Interconnected data centers and experimental facilities, Internet TCP/IP as common protocol, remote access	Distributed interconnected computing facilities, SOA and webservices, Grid as cooperative and distributed computing	Cloud adoption, infrastructure services on-demand Federated facilities and network access, Federated access and Identity management, 3G->4G	Distributed scalable computing, cloud based Big Data technologies, high performance networks, 5G technologies, wireless access, IoT sensor networks	Composable virtualized RI provisioning on demand, common federated computing and networking platform/environment, Cloud, DevOps and AI enabled, Digital Twins
Data	Proprietary formats, system or experiment specific	Standard format for data exchange, proprietary metadata	Domain/RI based data/metadata interoperability, custom data models, distributed storage, directories	Interoperable data, domain based metadata	FAIR data, Data Factories, Metadata registries, Interoperable/common Data Management model	Fully adopted FAIR principles, Semantically enabled scientific data lakes, secure/trusted data exchange, full data value chain
Infrastructure Management Technologies	Local management	Local management, management information exchange	Common Management Model, Distributed management, 3G Roaming	OSS/BSS, Automated deployment, adaptation, monitoring	Integrated Operation and Automation, Automated identity provisioning	Fully automated RI and services provisioning, management and operation, optimisation

EOSC Interoperability Framework (Feb 5, 2021)

- Defines 4 layers
  - Technical
  - Semantic
  - Organisational
  - Legal



https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-194348068

EOSC Interoperability Framework Feb 5, 2021

Layer	Recommendation
Technical (services but not yet infra)	• Open Specifications for EOSC Services. • A common security and privacy framework (including Authorisation and Authentication Infrastructure). • Easy-to-understand Service-Level Agreements for all EOSC resource providers. • Easy access to data sources available in different formats. • Coarse-grained and fine-grained dataset (and other research object) search tools. • A clear EOSC PID policy.
Semantic (Metadata)	• Clear and precise, publicly-available definitions for all concepts, metadata and data schemas. • Semantic artefacts preferably with open licenses. • Associated documentation for semantic artefacts. • Repositories of semantic artefacts, rules with a clear governance framework. • A minimum metadata model (and crosswalks) to ease discovery over existing federated research data and metadata. • Extensibility options to allow for disciplinary metadata. • Clear protocols and building blocks for the federation/harvesting of semantic artefacts catalogues.

EOSC Interoperability Framework Feb 5, 2021

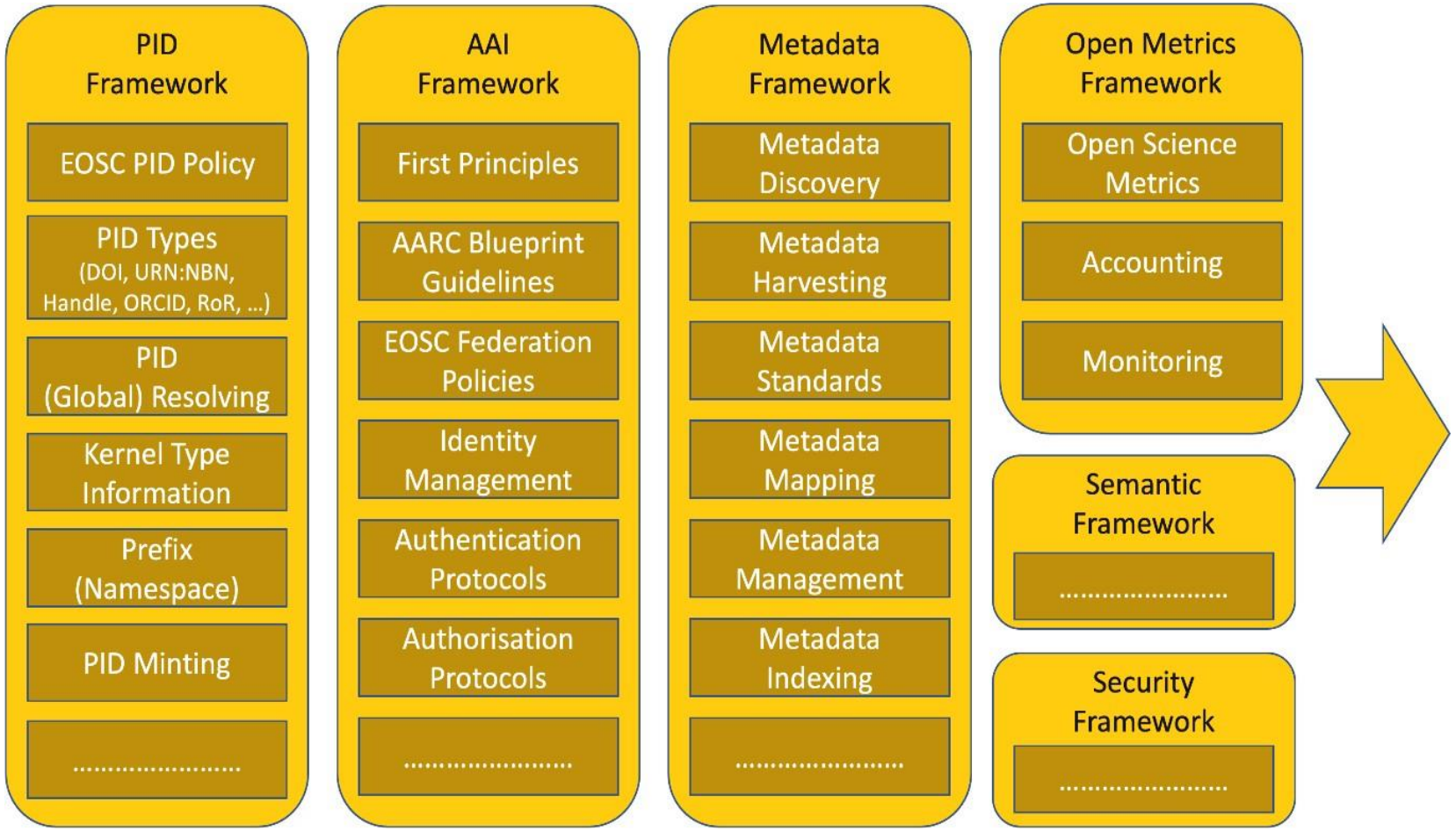
Layer	Recommendation
Organisational	• Interoperability-focused rules of participation recommendations. • Usage recommendations of standardised data formats and/or vocabularies, and with their corresponding metadata. • A clear management of permanent organisation names and functions.
Legal	• Standardised human and machine-readable licenses, with a centralised source of knowledge and support on copyright and licenses. • Permissive licenses for metadata (and preferably for data, whenever possible). And CC0 preferred over CC BY 4.0. • Identification of different parts of a dataset with different licenses. • Clearly marked instances of expired or non-existent copyright, as well as for orphan data. • EOSC-recommended licenses and their compatibility with Member States' recommended licenses. • Tracking of license evolution over time for datasets. • Harmonised policy and guidance to dealing with cases where patent filing or trade secrets may be compromised by disclosure. • GDPR-compliance for personal data. • Additional restrictions on access and use of data only applied in cases of applicable legislation or legitimate reasons. • Harmonised terms of use across repositories • Alignment between Member States national legislations and EOSC.

EOSC-IF Components

- Authentication and Authorization Interoperability (AAI) framework
- PID Framework
- Metadata framework
  - Resource and Provider Metadata profiles
- Data access framework
- Service management and access framework
- Open metrics framework
- Security framework
- Support framework

- Needs translation from “EOSC” language to technical language
- Still no clear design approach
  - Danger of silos and re-inventing
- No infrastructure vision

EOSC-IF: Structured View (EOSC Iron Lady)



- Technical terms to be clarified

EOSC-IF is about Data – To support FAIR

- FAIR Digital Object (FDO) is a key concept
- Technical Interoperability:
  - Artefact Common Protocols and Data Formats
- Semantic Interoperability:
  - Contextual Semantics related to Common Semantic resources
- Organisational Interoperability:
  - Contextual Semantics related to Common process resources
- Legal Interoperability:
  - Contextual licenses related to Common Licenses resources

- FDO is actively promoted by GO FAIR Technical Center and Peter Wittenburg
  - Recent presentation at e-IRG meeting (e-Infrastructure Reflection Group - EC policy consulting body)

FAIR is an Overloaded Concept (and a term)

- Primarily, FAIR is (set of) principles for sustainable Research Data Management (RDM) and Open Science
  - Findable – Accessible – Interoperable – Reusable
- FAIR is an initiative
- FAIR is a key policy area of EOSC
- FAIR data management is part of Data Management Plan (DMP) and required by Horizon Europe and many national funding bodies
- FAIR impose a number of requirements to Research Infrastructure
- Existing RIs run dedicated projects on FAIR adoption: ENVRI-FAIR, ELIXIR

- Universities should play important role in FAIR and RDM adoption
  - Still slow adoption at all levels: Bachelor, Master, Doctoral, teachers

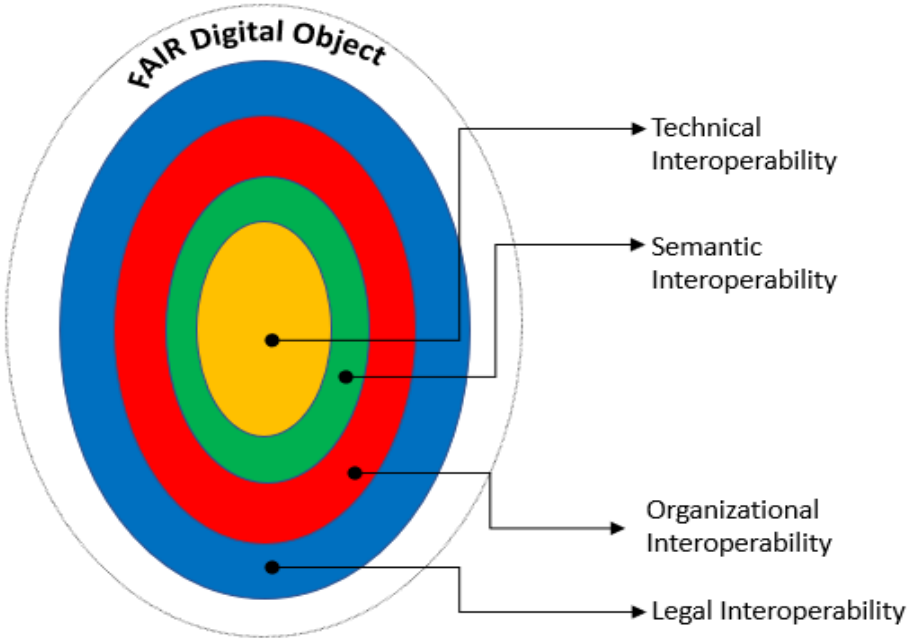
FAIR from the technical/infrastructure point of view

- **Findable**
  - Metadata and PID – infrastructure and tools
  - Registries and handles resolution, API
  - Policies and SLA
- **Accessible**
  - Repositories and data storage: infrastructure and management
  - Policy and access control: infrastructure and API management
  - Data access protocols
  - Usage Policy and Sovereignty
  - Data protection, compliance, privacy and GDPR
- **Interoperable**
  - Standard data formats
  - Metadata and API
  - FAIR maturity level and certification
- **Reusable**
  - Data provenance and lineage
  - Preservation, archiving
  - Metadata, PID and API – linked or embedded into datasets

New profession of the **Data Steward** is promoted by EOSC an currently demanded by industry. Main competences include:

- Data management
- Domain research or business knowledge

FAIR Digital Object – A core for EOSC-IF



- In EOSC, a digital object can be research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc
- FAIR Digital Object Extends Digital Object concept for better FAIRness
- FAIR Digital Object (FDO) is a core building block of EOSC-IF
  - Four interoperability layers applied
  - Requires infrastructure support

FDO (FAIR Digital Object) and PID Infrastructure Requirements

- General requirements include **machine actionability**, technology independence, **persistent binding,** abstraction and structured hierarchical encapsulation, compliance with standards and community policies (as specified in the FDO general requirements G3-G9);
- ***FDO is identified by PID***; there are possible multiple PID frameworks defined by PDI scheme, namespaces, ontologies or controlled vocabularies (FDOF1);
- A **PID resolves to a structured record (PID record)** with attributes that are semantically defined within a (data) type ontology (which may be defined for different application or science domains) (FDOF2);
- PID record may include other attributes that are important to characterize specific types of FDO or that are required by applications. Additional attributes must be registered in a ***data types registry*** (FDOF4);
- Metadata used to describe FDO properties should use standard semantics and ***registered schemes*** to allow machine readability and actionability (FDOF8-FDOF10).

- FAIR Digital Object Framework, Technical implementation guideline, version 1.02, [online] https://datashare.rzg.mpg.de/s/RTeYZGe3QMgEciH/download?path=%2F&files=FAIR%20Digital%20Object%20Framework-v1-02.pdf

Discussion

- **FAIR and Open Science implementation in organisations**
- What do you think is the most important challenge in achieving FAIRness
- **Challenges with technical interoperability SLICES and EOSC**
- Challenges with semantic interoperability SLICES and EOSC

- Go to **www.menti.com** and use the code **4309 9563**

Mentimeter: FAIR and Open Science implementation in organisations – "Maturity Level Indicators"

- FAIR awareness: Researchers and RI operators
- Top management aware and organization has a plan to implement FAIR
- Organisation has approved DMP including FAIR requirements
- Organisation has DMP, Data Governance Policy and Data Steward or Data Curator
- Open Data Policy accepted and research data are mandatory published in open access repositories
- Metadata catalog and data publishing tools used
- Metadata profile defined for discipline/domain data
- Metadata are added at the initial stage of data collection (from experiment)
- To ensure Reusability, data are published with API and application
- Use of existing data evaluated: Data search and review at the initial stage of the project

Mentimeter: What do you think is the most important challenge in achieving FAIRness

- Cast your short few words reflection

Technical Interoperability with EOSC Challenges -> Mentimeter

(i) fAir: Separate authentication and authorization are often required when accessing services across different infrastructures and communities, which generally requires transfer of personal information among identity and service providers. To address this issue, there is a need to develop an AAI framework that is community independent and minimally obstructive;

(ii) faIr: Research data may be stored in different formats which are either general purpose (CSV, XML, JSON, etc.) or community specific (Darwin core, FITS, VOTable, VOResource, etc.) which are difficult to reuse across communities. To solve this, a common-minimum metadata model is needed to allow seamless discovery and reuse of data across multiple formats;

(iii) FaiR: Research data is often not available in multiple granularities. This makes it difficult to be found and reused by different scientific domains requiring different granularity of data. Thus, there is a requirement for research data to be stored at multiple levels of abstractions (fine grain and coarse grain) so that a wide variety of scientific and application domains can benefit from it;

(iv) PID: Generally, scientific communities employ community-specific persistent identifiers (PURL, IUPAC international chemical identifier, DOI, etc.) with a different set of policies. This sometimes results in identifiers which can be difficult to resolve. Therefore, a community-agnostic PID policy is required for a common understanding.

Semantic Interoperability Challenges -> Mentimeter

- **Semantic interoperability** refers to the ability that the exchanged data is understood well and have a common meaning across different entities of the EOSC ecosystem.

(i) Semantic artefacts are poorly documented and definitions of terms used are not precisely defined. This makes it difficult to be used across communities;

(ii) Furthermore, common reference repositories/registries for semantic artifacts are not easily available or maintained for long enough;

(iii) Lack of common metadata schemas across communities. Different communities use different metadata schemas such as DarwinCore, RDA metadata, DCAT, DDI4, etc.

EOSC Reference Information

- EOSC Minimal Viable EOSC and EOSC Core
- EOSC Technical Architecture by EOSC-hub

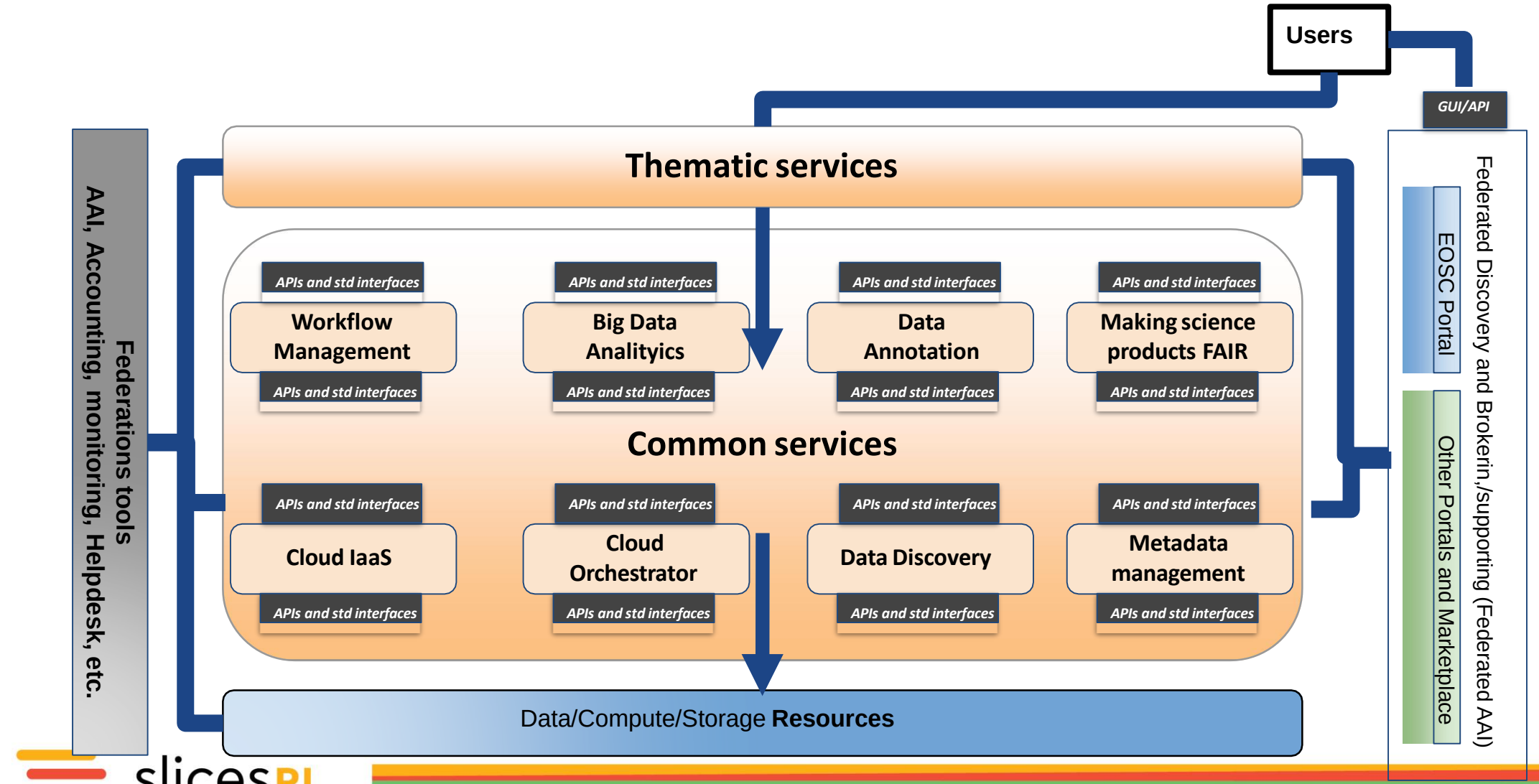
https://www.eosc.eu/task-force-faq

- **Who are the members of the Task Forces?**
  This is still to be defined but will be agreed by September 2021. Much depends on the rules defined by each Task Force. Some will restrict the size to 25-30 members whilst others will not set a limit in membership. The Board will select the members in collaboration with the charter drafting coordinators.
  - The open call for drafters of the Task Forces in March 2021 resulted in 255 applications.
  - The Board nominated four coordinators per Task Force to steer the work and assigned applicants to a single Task Force ensuring balance in expertise, preference of applicants for specific Task Forces, gender, and organisations.
- **Do I need o formally apply if the membership of a Task Force is open?**
  Yes. You need to apply so that all members of the Task Force match the skillset and selection criteria for that Task Force and so that we have a record of the names and contact details of all members of the Task Forces.

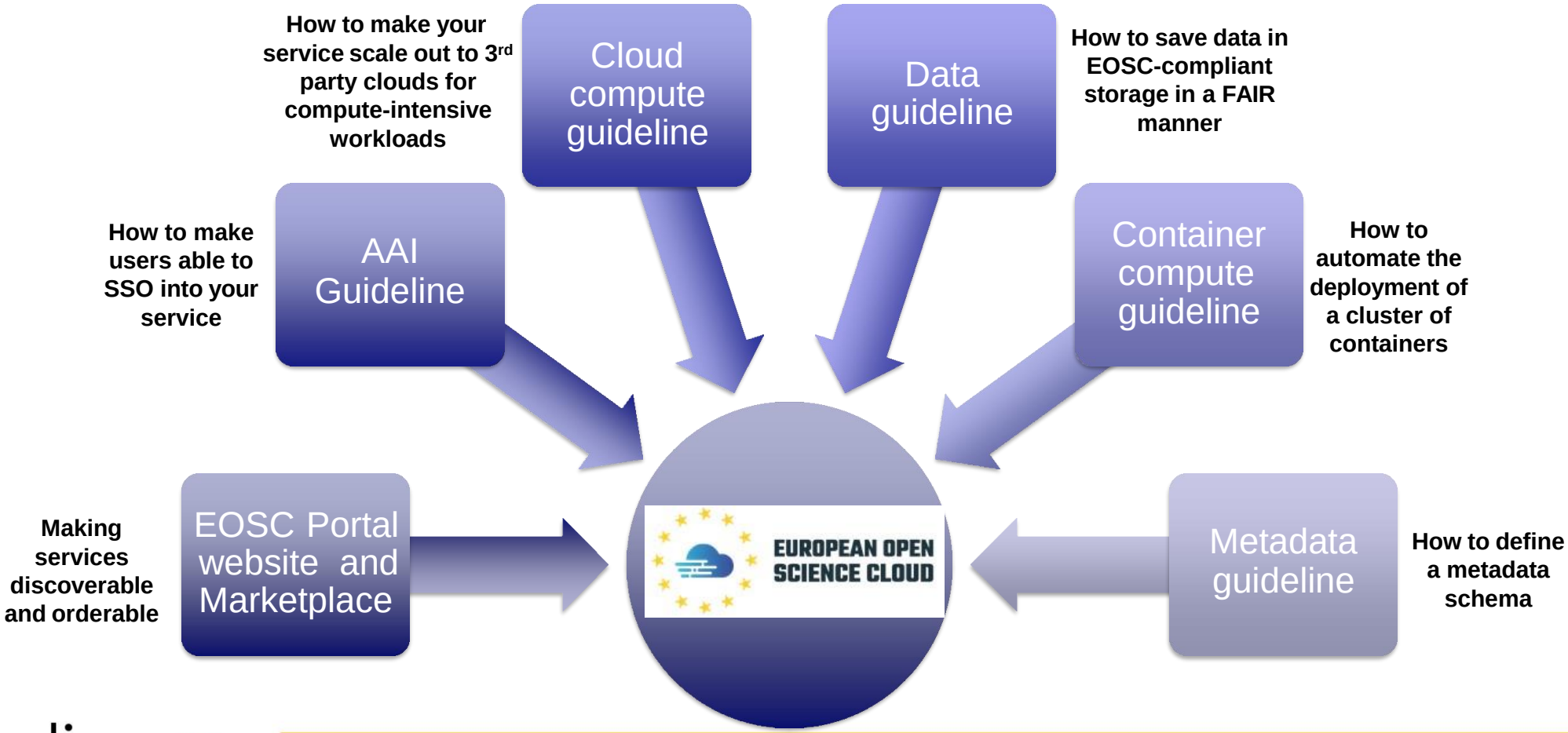
- **Will there be funding for activities in the Task Forces?**
  There is no budget foreseen in the first months for conducting studies or supporting activities in the Task Forces.

EOSC Technical Architecture

https://documents.egi.eu/public/ShowDocument?docid=3495

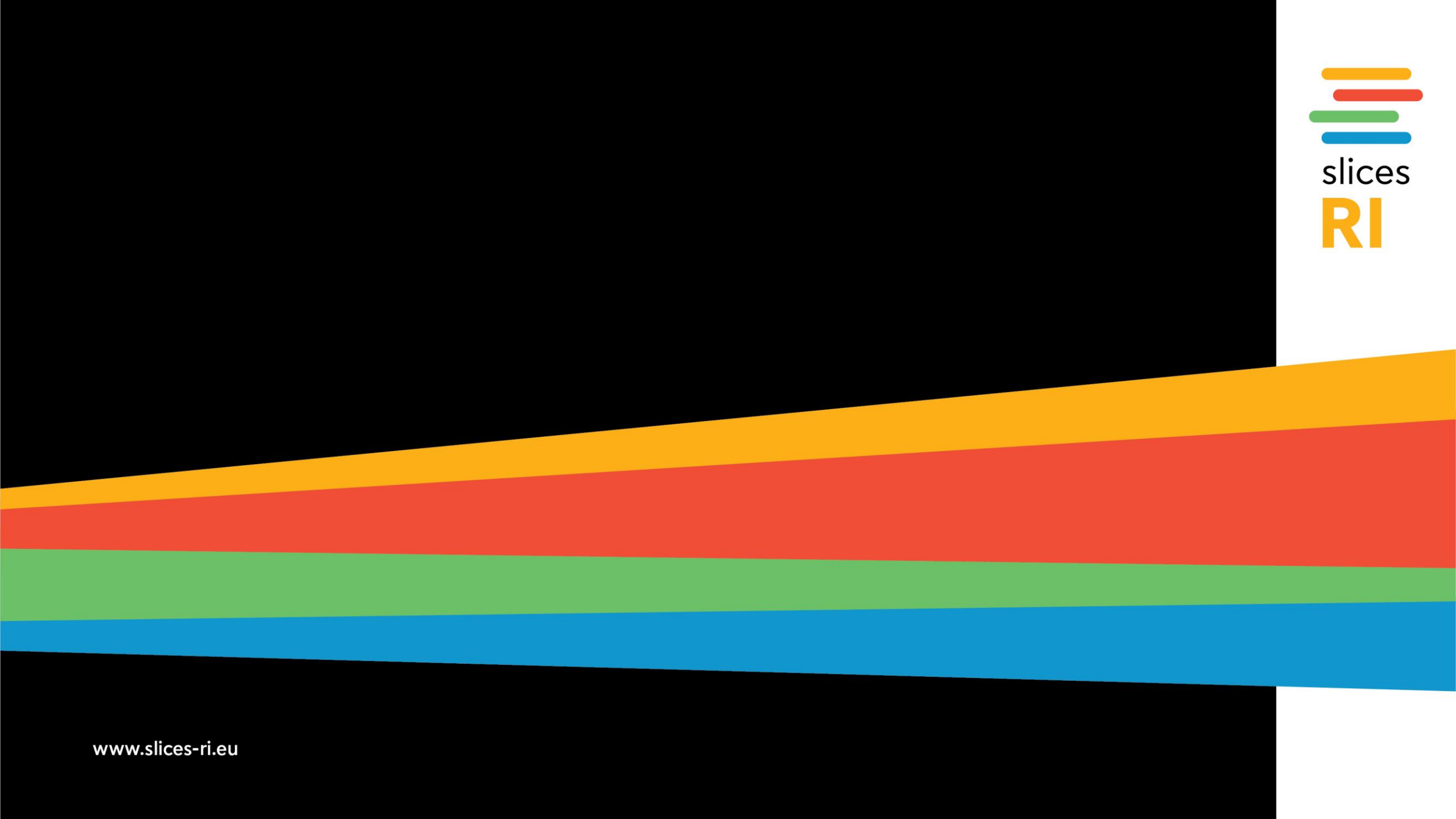


EOSChub Reference architecture, interfaces and interoperability guidelines



From EOSC-1 to EOSC-2: Four Technology Aspects

RI Type	EOSC-1 (2016-2018)	EOSC-2 (future) (2020-2025)	Beyond 2025	
Definition	Interoperable Federated (European) RI, FAIR RI	Virtualised Pan-European **RI platform as a Service and Ecosystem (PRIaaS)**		
Network & Compute	• Distributed scalable computing • Cloud based Big Data technologies • High performance networks • 5G technologies, wireless access • IoT sensor networks • **Portal and Services Catalog** • Industry standards and IDSA adoption	• Composable virtualized RI provisioning on demand (including for services integration) • Common federated computing and networking platform/environment, enabling virtual RIs • Cloud based and cloud enabled • DevOps and AI enabled services • Digital Twins • Interoperability and Integration with Industry infrastructure (e.g. IDSA+, Industrial Internet)		
Data Infra	• FAIR data • Data Factories and PID • Metadata registries • Interoperable/common Data Management model	• Fully adopted FAIR principles, extended to ontologies • Semantically enabled scientific data lakes, common vocabularies • Secure/trusted data exchange (data markets) • Full data value chain supported (cross-domain)		
Security	• Federated Identity Management, Federated Access Control • Automated identity provisioning	• Zero trust security, Trust Bootstrapping • Homomorphic encryption and data processing • Quantum ready encryption, Quantum enabled key management • Federated Identity Management, Federated Access Control • Automated identity provisioning		
Infra Managnt Technolog	• Integrated Operation and Automation	• Fully automated RI and services provisioning, management and operation • Optimisation of infrastructure and operation • DevOps and AI enabled (re-usable design patterns)		


slices
RI
www.slices-ri.eu