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)

www.slices-ds.eu

Executive Summary

This document describes the SLICES Interoperability Framework (SLICES-IF), which main objective is to enable the exchange of research data with the European Open Science Cloud (EOSC) and facilitating the federation of resources with other Research Infrastructures (RIs). In addition, SLICES is committed to achieve the Open Science goals defined by the European Commission (EC) by promoting the Open Access of research data, tools, software and infrastructure design principles. The EOSC being a key pillar of implementing Open science paradigm, interoperability with EOSC at all the levels of technical, semantic, organizational and legal interoperability will establish SLICES as a key proponent of Open science. The first step in achieving interoperability with EOSC is to fully understand and adopt the key technical and sematic design choices recommended by EOSC interoperability framework (EOSC-IF). In this document we have thoroughly analysed the technical and semantic aspects of the EOSC-IF which are used to define the SLICES-IF. Likewise, the EOSC-IF, the FAIR digital object (FDO) guidelines are at the core of SLICES-IF to enable the smooth research data exchange with EOSC and ensure the transparency, discoverability, reusability and reproducibility of the scientific resources, experimental workflow and the generated knowledge.

By defining the SLICES-IF, we have provided the design choices that are needed to integrate SLICES with EOSC. This will allow the SLICES users to benefit from the resource-rich EOSC ecosystem having a plethora of scientific tools, software, research data and infrastructure resources and thereby helping to realize the European vision of 'shared research knowledge system'. We start by defining the key requirements for SLICES-IF emanating from the SLICES use cases and the proposed integration with EOSC and other European RIs. We then provide the choices made for the key semantic and technical building blocks such as Authentication and Authorization (AAI), persistent identifiers (PID) policy and application programming interfaces (APIs). These design choices are selected in such a way that they fulfil the interoperability requirements. For metadata model and profiles, deliverable D4.3 titled 'SLICES metadata profiles to support FAIR principle' is dedicated and hence not considered here.

Table of contents

1. Introduction

SLICES aims to develop a pan-European digital research infrastructure (RI) for data-driven sciences. In data driven scientific research, computing, storage and networking play important role. SLICES is committed to push forward the state-of-the-art in digital RIs by developing new computing and networking architectures and protocols as well as develop tools for better accessing, discovering and monitoring digital RIs. One of the coming steps for SLICES is to connect the nodes among SLICES-RI partners and ensure their integration with European Open Science Cloud (EOSC)[1]. This will accelerate the digital sciences by providing access to unique experimental facilities and testbeds and support experimentation with a sufficient amount of storage and processing power for large scale scientific experiments such as artificial intelligence and its applications in various domains of scientific research.

To achieve excellence in scientific research and innovation, Open Science[2] is being seen as one of the most important policy directives proposed by the European Commission. The concept of open science includes open access to research outputs (publications, data, models, software, tools, etc.), clear data management plan with adherence to FAIR practices in data representation and storage, ensuring reproducibility of research results and encouraging participation of various stakeholders including citizens in the scientific process. The recent report[3] from the European Commission Open Science Policy Platform (OSPP) advocates for the 'shared research knowledge system'. The core emphasis of the 'shared research knowledge system' is the systemic effort to foster all practices and processes that enable the creation, contribution, discovery and reuse of research knowledge more reliably, effectively and equitably. The trustworthiness, integrity and transparency in research systems are mentioned as the core values for fostering responsible innovation and scientific excellence. To this end, SLICES is committed to adopt the principles of 'open science' and 'shared research knowledge system' by: (i) implementing a pan-European federation of digital RI using open specifications for implementation of different building blocks to accelerate the large-scale scientific collaborations; and (ii) necessitating the adoption of FAIR principles in storage and distribution of research data, publications, software and infrastructure to achieve reusability, transparency, reproducibility and knowledge sharing.

EOSC has established itself as an important pillar in the implementation of Open science concept by accelerating the adoption of the FAIR data practices among researchers in the European Union. Integration of SLICES into European Research Infrastructure via EOSC will facilitate data sharing and reuse among SLICES partners and the larger European researchers' community. This naturally aligns SLICES to the concept of Open science. Interoperability-focused Integration of SLICES with EOSC will make it easier for SLICES users to take the benefits of a large number of services and tools pertaining to diverse scientific domains that are being developed around the EOSC ecosystem. As a result, SLICES will help to realize the vision of 'shared research knowledge system' proposed by the EC.

The main objective of D4.2 is to design the integration framework of SLICES with EOSC in such a way that the data exchange between SLICES and EOSC is interoperable for scientific workflow management for data storage, processing and reuse. To this end, the recommendations of the EOSC interoperability framework[4] are studied in detail and based upon which the SLICES interoperability framework is proposed. In this deliverable, we mainly focus on the technical and semantic

---

[1] EOSC portal, https://eosc-portal.eu/, [Last accessed 27 August 2021]

[2] Open Science webpage, https://openscience.eu/, [Last accessed 27 August 2021]

[3] Progress on Open Science: Towards a Shared Research Knowledge System Final Report of the Open Science Policy Platform, European Commission, 2020

[4] EOSC Interoperability Framework [Last Accessed 15 July 2021], https://op.europa.eu/en/publication-detail/-/publication/d787ea54-6a87-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-190308283, [Last accessed 27 August 2021]

interoperability aspects while the legal and organization interoperability is considered in the deliverable D3.1 titled 'Governance structure and possible types of legal entities' in the section about the Data Governance Framework. In the context of technical and semantic interoperability, the major building blocks such as metadata model, semantic artifacts, authentication and authorization interoperability (AAI), persistent resource identifiers (PID) and application programming interfaces (APIs) are analysed and design recommendations for SLICES interoperability framework are proposed.

The remaining of this document has been organized as the following. Section 2 presents a brief discussion of SLICES use cases by focusing on why the SLICES interoperability framework is required for data exchange among SLICES resources, EOSC and other research infrastructures. In section 3, the EOSC interoperability framework is analysed and key takeaways from the EOSC-IF, which becomes basis for defining the SLICES-IF, are highlighted. Section 4 provides a detailed discussion on the technical and semantic requirements of SLICES-IF and defines the key principles of the SLICES-IF aligned with the EOSC-IF and FAIR digital object guidelines. Section 5 discusses the key components of the SLICES-IF such as AAI, APIs and PID frameworks. SLICES Metadata model and profiles are part of the deliverables D4.3 and hence excluded in this document. In Section 6, Open Science concept of Horizon Europe is discussed in the context of SLICES. Finally, section 7 provides a general discussion about this deliverable.

2. SLICES use cases Overview

SLICES aspires to provide a fully programmable and remotely accessible digital research infrastructure comprising of geographically distributed compute, storage and networking (wired and wireless) elements. From the use case perspective, SLICES would primarily enable large scale experimental research the following three categories:

- <u>Distributed cloud/edge computing</u>: under this category, experiments regarding infrastructure operation and management are considered (e.g., management of virtualized resources), as well as applications of Machine Learning and Artificial Intelligence that can run in a distributed manner (e.g., Federated Learning).

- <u>IoT verticals</u>: under this category of experiments regarding applications and ubiquitous IoT networking, interoperability of sensor infrastructure and integration of IoT infrastructure with the rest of the stack are envisioned to take place.

- <u>Wired/ wireless networking:</u> experiments regarding prototyping next generation protocols for network interconnection (e.g., researching new waveforms and new frequencies, spectrum and wireless network management, formation of Heterogeneous networks) are foreseen under this category.

These three categories can be combined in a single use case, for instance. Indeed, the SLICES Research Infrastructure can be used to test and validate the integration and deployment of IoT sensors in the context of a smart city. Typically, these Internet of Things sensors are basically wireless sensors monitoring specific environmental parameters such as the noise level, the air quality, the temperature, etc. Each IoT sensor is connected to a wireless sensor network (WSN). Different WSNs could be installed in the smart city, depending on the wireless communication protocols. Bridges and gateways permit access to the different WSNs from the cloud in a distributed architecture where the data are collected and centralized to the cloud. Some local processing activities and storage can be done at the edge with the data generated by the sensors to avoid a lot of transmissions between the edge and the cloud. This kind of deployment can be tested currently in a research infrastructure such

as Fed4FIRE+ testbeds at a reasonable scale, in particular to validate the integration and the interoperability of the heterogeneous data sources. Through the SLICES Research Infrastructure, the experiment of such deployment will be facilitated, notably by the features offered by SLICES-RI in terms of large-scale experimentation in a distributed environment, will be offered the possibilities of more resources and more diverse technological specs, and will be enriched by other aspects notably as follows. This use case also permits the collection of data that can be published and stored in the EOSC marketplace. Indeed, the data collected in the context of the experiment can be reused in other experiments associated with smart cities or for other IoT verticals like smart water, smart building or smart mobility/transport. The correlation of different sources of data could lead to a better understanding of the behaviour of the smart city, notably by a cross-domain approach based on different involved IoT verticals.

This use case implies some functional requirements for the publication of the generated data. Mainly, the FAIR principles should be fully respected during the process of data publication on the EOSC marketplace, notably for the metadata. The EOSC marketplace can also serve to retrieve data from other research infrastructures similar to Fed4FIRE+ and to correlate the external heterogeneous data with the data produced in the SLICES Research Infrastructure during an experiment. The role of EOSC is also to be an interoperability point at high-level to exchange data and knowledge acquired during similar experiments. **To summarize, the SLICES use cases would require the interaction among different SLICES infrastructure components and the data exchange with EOSC. This motivates the need for a clear and precise interoperability framework for SLICES** as explained:

i)   SLICES aims to build a distributed pan-European research infrastructure that shall be accessible remotely. Accessing different tools and research data that is distributed in nature would need different SLICES components (research data, computing and storage modules, providers and users) to communicate with each other. To provide a seamless interaction of these components to enable a coherent research environment, interoperability is a must;

ii)   In order to comply with the FAIR principles, interoperability is essential to ensure that the SLICES data can be integrated with scientific workflows for analysis, storage and processing;

iii)   Interaction with EOSC and other RIs is an important objective for SLICES to consolidate the scattered European research infrastructure. SLICES-IF is essential to facilitate this.

3.   Recommendations/analysis of EOSC interoperability framework

Interoperability is an essential feature of EOSC ecosystem as a federation of services and data exchange is unthinkable without interoperability among different EOSC constituents. The meaningful exchange and consumption of digital objects is necessary to generate value from EOSC which can only be realized if different components of the EOSC ecosystem (software/machines and humans) have a common understanding of how to interpret and exchange them, what are the legal restrictions, and what processes are involved in distribution, consumption and production of them. To facilitate this, EOSC interoperability framework (EOSC-IF) is defined as a generic framework for all the entities involved in the development and deployment of EOSC. Instead of proposing specific recommendations, EOSC-IF mainly discusses the common requirements and challenges faced by the user communities targeted by EOSC, and potential solutions to meet these requirements to achieve interoperability at different levels.

3.1. EOSC interoperability challenges and requirements

EOSC defines four types of interoperability components, namely: technical, semantic, organizational and legal.

Technical interoperability is defined as the ability of different information technology systems and software applications to communicate with each other and seamlessly exchange data. The main challenges in the way of ensuring technical interoperability are:

(i) 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) 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) 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) 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 refers to the ability that the exchanged data is understood well and have a common meaning across different entities of the EOSC ecosystem. The major obstacles to ensure semantic interoperability are:

(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) Also, there is a lack of common metadata schemas across communities. Different communities use different metadata schemas such as DarwinCore, RDA metadata, DCAT, DDI4, etc.

All these problems (i)-(iii) bring ambiguity in deriving and discovering logic, inference and knowledge from the shared data. To address these challenges, principled approaches for the creation and maintenance of ontologies and metadata schemas are required. Further, the metadata schemas should be extensible to allow for domain-specific attributes for harmonization across different scientific domains. It is important to mention that not only for research data but also the extensions should be available for scientific workflow, methods, software, hardware and experimental facility and laboratory protocols to facilitate a truly domain-agnostic semantic interoperability.

Organizational interoperability is focused on the alignment of organizational policies, functions, responsible people, documentations and processes across different EOSC service providers. The main

emphasis is on defining a governance framework to achieve cross-organizations and cross-discipline interoperability. The main challenges include:

(i)   There is a lack of clear description of the 'terms and conditions' and `acceptable use policies' that must be adhered by services provisioned by EOSC;

(ii)   Rules of participation do not provide details of how interoperability will be achieved across organizations and domains;

(iii)   Sustainability policy (long-term availability) of services and infrastructures is not always available. Therefore, there is a need for defining a clear governance framework concerning the different functions and policies for participation in EOSC, documentation defining unambiguous terms and conditions and acceptable use policies are needed. Furthermore, interoperability certifications for service providers need to be developed so that users are aware of the interoperability levels of different services.

Legal interoperability primarily concerns data access governed by various forms of intellectual property rights (e.g., licensing, copyrights, etc.), general data protection regulation (GDPR), private and sensitive data and enabling legal instruments. Ensuring legal interoperability is very challenging due to the following:

(i)   Data reuse is difficult without clear information about the rights and legal conditions for reusing the data;

(ii)   Different datasets may have different licenses not compatible with each, thus making it difficult to combine and use them together;

(iii)   In case of some old datasets, it is difficult to trace the copyright holder due to missing information which makes them difficult to reuse;

(iv)   The scope of national copyrights may vary across jurisdictions, making it difficult for users to reuse the data;

(v)   Separate licensing for different embedded objects in a dataset (e.g., photos in a dataset may have different licenses) can be confusing;

(vi)   Users' rights for using the data (e.g., commercial use) may change with the passage of time;

(vii)   GDPR introduces strict constraints on sharing and processing of personal and sensitive data resulting in disproportionate costs on safeguarding personal data and obtaining individual consent for individual datasets;

(viii) Each EU member state has a different guideline for data privacy impact assessment (DPIA)of high-risk personal data. In the next section, we summarize the key EOSC recommendations pertaining to the above four interoperability criteria.

3.2. EOSC interoperability recommendations

Technical: (i) EOSC recommends that all the services should be using open specifications whenever possible; (ii) A common security and privacy framework including a common authentication, authorization mechanism should be used; (iii) A clear policy for persistent identifiers (PID) for research data, infrastructure and software should be defined; and (iv) data should be made available in a different format for the ease of accessibility.

Semantic: (i) All the concepts, metadata and schemas should use clear, precise and publicly available definitions which are referenced with PIDs; (ii) a minimum metadata model should be used to describe all the research data for ease of discovery; (iii) metadata should have extensibility options for the inclusion of domain specific information; and (iv) semantic artifacts should contain the associated documentation.

Organizational: Rule of participation should be clear for the resource/service providers with well-defined management functions.

Legal: (i) compatibility of EOSC licences with member state licenses should be ensured and there should be a clear alignment of the member state legislations with EOSC legislations; (ii) GDPR compliance of personal data should be adhered; and (iv) policy and guidelines w.r.t. patent filing, trade secret disclosure and data access restriction should be harmonized across participating member states.

3.3. EOSC FAIR Digital Object and PID framework

3.3.1. FAIR digital framework in EOSC

FAIR Digital Object (FDO) is the core building block of EOSC-IF. Here, Digital Object refers to the kind of objects that allow binding all critical information about any entity. In EOSC, a digital object can be research data, software, scientific workflows, hardware designs, protocols, provenance logs, publications, presentations, etc., as well as all their metadata (for the complete object and for its constituents). Figure 1 shows a schematic of FDO. An FDO should conform to all the four layers of interoperability introduced earlier in this document by following the FAIR guidelines.
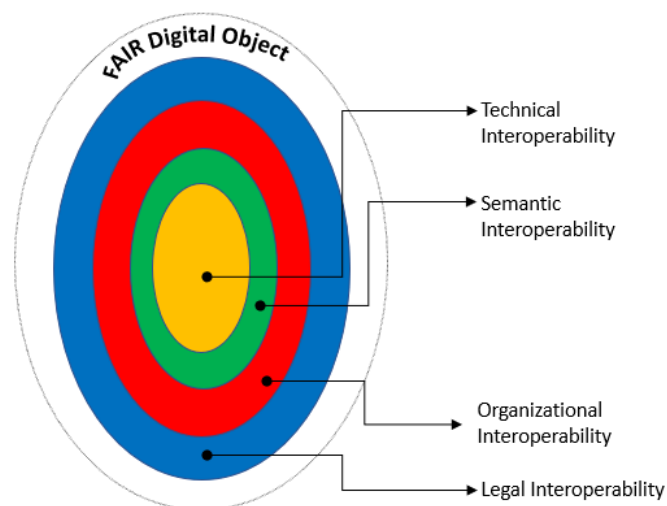
Figure 1 – EOSC FAIR Digital Object illustration

3.3.2. FAIR Digital Object definition by FDO Forum

FDO (FAIR Digital Object) and **PID (Persistent Identifier)** are key components of the EOSC architecture and supporting federated data infrastructure that ensures consistent implementation of the FAIR data principles. FAIR principles are realized through FAIR Digital object, and PID services and infrastructure provide facilities to access and manipulate FDO.

According to FDO Forum, the FAIR Digital Object concept brings together FAIR guiding principles and Digital Object which supports interoperability across existing and evolving data regimes/frameworks using mechanisms of the structured machine-readable identifiers and principles of persistent binding for digital object of different types. It is important to mention that activities currently coordinated by FDO Forum are including essential results and current activities at the RDA (Research Data Alliance)

on Data Types and Data Types Registries, Data Factories, others, and are aligned with EOSC technical development what is reflected in the EOSC Architecture and EOSC Interoperability Framework.

FDO Forum provides the following definition of the FDO[5]:

"A FAIR digital object is a unit composed of data and/or metadata regulated by structures or schemes, and with an assigned globally unique and persistent identifier (PID), which is findable, accessible, interoperable and reusable both by humans and computers for the reliable interpretation and processing of the data represented by the object."

The following main requirements to FDO services and infrastructure defined in the FDO Framework[6] (note, below list provides in brackets requirement IDs as they are identified in the cited document):

- 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);

- The structured PID record includes at least a reference to the location(s) where the FDO content and the type can be accessed, and also metadata describing FDO can be retrieved (FDOF3);

- 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);

- Each FDO is accessed via API by specifying PID and additionally point. API must support basic operation with FDO: Create, Read, Update, Delete (referred to as CRUD), however subject to access control and policy (FDOF5, FDOF6);

- Metadata used to describe FDO properties should use standard semantics and registered schemes to allow machine readability and actionability (FDOF8-FDOF10).

Such FDO definition aims to specify a core model for interoperable data that meet the FAIR guiding principles for research data and vision of the future "Internet of FAIR data services" when combined with PID services and infrastructure. When developing the FDO/PID framework, the developer community used similarity and principles of the Internet in implementing and managing IP-address system (similar to PID) and Internet Domain Name Systems (DNS) as an inspiration for PDI and FDO properties resolution and discovery.

3.4. EOSC PID Architecture and PID Policy

PID Framework and corresponding infrastructure are recognized as key components to implementing Open Science and European Open Science Cloud into reality. There are two documents defining PID framework for EOSC published by European Commission in consultation with the EOSC community:

---

[5] Base and full technical definitions of a FAIR Digital Object (FDO), [online] https://datashare.rzg.mpg.de/s/RTeYZGe3QMgEciH/download?path=%2F&files=FDO-Definition-15Apr2021.pdf, [Last accessed 27 August 2021]

[6] 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, [Last accessed 27 August 2021]

- PID Architecture for EOSC[7];

- Persistent Identifier Policy (PID) for the European Open Science Cloud[8]

The PID technical architecture is defined to provide a direction to creating a common interoperable PID and FDO infrastructure and services in EOSC as well as provide a basis for interoperability between existing PID frameworks/systems that are created by sectoral sciences or developed globally by the Internet community.

PID systems are defined as technically implemented services to generate (also referred to as mint) and resolve PIDs. PID systems are the technical hardware and software components that provide PID related services where PID framework defines a governance structure together with a technical infrastructure that allows organizations to implement PID services for their needs/domain and within agreed policy.

Typically, PID is perceived as a string used as a reference. This string usually has internal syntax to reflect the infrastructure used to access digital or data objects. Similar to Internet or web addresses and URL, PID typically includes location address, namespace definition, and a set of attributes needed to access DO and maintain links between related/linked objects. PID must be unique and precisely refer (point or direct) to an individual object. Such PID structure of interconnected digital objects is defined as PID Graph being developed in implemented in EOSC by the FREYA Project[9]. The importance of creating a consistent PID infrastructure (prospective PID Graph) is recognized in the new Horizon Europe that defined a PID implementation as a priority topic in HE2021-2022 Calls HORIZON-INFRA-2021-EOSC-01-03: Deploying EOSC-Core components for FAIR[10].

PID Architecture and PID policy documents define the core roles/actors, their responsibilities and interaction with the core PID services. Similar to Internet and DNS, the following services are defined for PID architecture: PDO repository (typically PID enabled storage), metadata repository, registry, resolver, type registry, local and global PID infrastructure.

Main roles include: PID authority, PID Service Provider, PID Manager, PID Owner (actually referring to the digital object owner), End User or Consumer (can both a person and a machine or service).

The PID Policy document implies requirements to the PID Provider and services to be at Technology Readiness Level not less than 9 (actual system proven in operational environment).

3.5. Existing PID Frameworks

There are few existing PID frameworks that use different approaches to express and construct digital object identifiers and operate infrastructure for digital objects discovery and search (that in general operates similar to DNS system or as a global catalogue).

3.5.1. Handle based systems

Handle based systems use the URL format for inserting the PID string that is separated by the slash sign "/" where the handle namespace is managed by the Digital Object Numbering Authority (DONA), responsible as registration authority and also the registrar for the top-level namespaces that are delimited by the by the left most digits to the first dot ".". These top-level namespaces are delegated

---

[7] PID Architecture for the EOSC, European Commission, [online]
https://op.europa.eu/en/publication-detail/-/publication/3136c3e6-4f07-11eb-b59f-01aa75ed71a1, [Last accessed 27 August 2021]
[8] Persistent Identifier (PID) policy for the European Open Science, European Commission, 15 Oct 2020 [online]
https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1, [Last accessed 27 August 2021]
[9] Integration of the PID Graph with the EOSC, DeliverableD4.5, FREYA Project [online] https://www.project-freya.eu/en/deliverables/freya_d4-5.pdf, [Last accessed 27 August 2021]
[10] Annex 3, Horizon Europe, WOrk Programme, Annex 3. Research Infrastructures [online]
https://sciencebusiness.net/sites/default/files/inline-files/Annex%203%20Research%20Infrastructures.pdf, [Last accessed 27 August 2021]

to the so-called Multi Primary Administrators (MPA) which are responsible for the whole prefix namespace starting with their digits of the top-level namespaces.

The following are examples of the handle-based systems:

- ePIC[11]: The ePIC Persistent Identifier Consortium for eResearch10 is an international consortium of currently nine members that signed a contract to provide a reliable Handle-based PID infrastructure for research data;
- B2Handle (from the EUDAT Service catalogue)[12]: EUDAT's B2HANDLE is a distributed service for minting, storing, managing and accessing persistent identifiers (PIDs) and essential metadata (PID records) as well as managing PID namespaces;
- DataCite[13]: DataCite is a leading global non-profit organisation that provides persistent identifiers for research data and other research outputs. Organisations within the research community join DataCite as members to be able to assign DOIs to all their research outputs. DataCite has more than 200 member organisations as of October 2020;
- DataCite is a member of the DOI Foundation[14] and uses DOIs as persistent identifiers. The DOI Foundation is a non-profit membership organisation with currently 10 DOI registration agencies. DOIs use the handle system for PID resolution, and this service is provided for all DOIs by CNRI[15], which provides a globally distributed network of about 50 proxy servers. The DOI Foundation is an MPA and member of DONA.

3.5.2. URN systems

The URN framework uses prefixes and namespaces to ensure the uniqueness of the PIDs and local PID services are provided to register/modify PIDs and resolve PIDs based on these prefixes. IANA is the central namespace authority for URN framework. It maintains the registry of URN namespaces at IANA[16]. The following are example of the URN based systems:

- URN:NBN: National Bibliography Number (NBN) is a generic term referring to a group of identifier systems administered by national libraries and institutions authorized by them. The identifier is described in RFC 845819;
- *URN:ISSN:* ISSN is an authoritative standard identifier system for continuing resources and in particular serial publications20. URN:ISSNs resolve to a metadata record describing the identified resource in the ISSN Portal;
- *URN:ISBN:* The ISBN (International Standard Book Number) is a unique machine-readable identifier, which identifies unambiguously any edition of a text-based monographic publication that is available to the public.

3.5.3. HTTP based systems

HTTP based identifier systems are represented by the following implementations:

- Cool URIs[17]: Used to publish information and relationships between them for the so-called Semantic Web the Resource Description Framework (RDF);
- URLs: A PURL[18] is a persistent URL, it is registered through the PURL service and provides a permanent address to access a resource on the web. When a user retrieves a PURL they will

---

[11] ePIC webage: The purpose of persistent identifiers (PID), https://www.pidconsortium.net/, [Last accessed 27 August 2021]

[12] EUDAT Collaborative Infrastructure: B2HANDLE, https://eudat.eu/services/userdoc/b2handle, [Last accessed 27 August 2021]

[13] DataCite website, https://datacite.org/, [Last accessed 27 August 2021]

[14] International DOI Foundation (IDF), https://www.doi.org/, [Last accessed 27 August 2021]

[15] Corporation for National Research Initiatives (CNRI), https://www.cnri.reston.va.us/, [Last accessed 27 August 2021]

[16] Formal Uniform Resource Names (URN) Namespaces, https://www.iana.org/assignments/urn-namespaces/urn-namespaces.xhtml, [Last accessed 27 August 2021]

[17] W3C: Cool URIs for the Semantic Web, https://www.w3.org/TR/cooluris/, [Last accessed 27 August 2021]

[18] PURL Administration: http://purl.org, [Last accessed 27 August 2021]

be redirected to the current location of the resource. When an author needs to move a page, they can update the PURL to point to the new location;

3.5.4. Intrinsic Identifiers

Intrinsic identifiers are represented by the widely used ORCiD[19] system that is a self-containing system that uses local/internal identifiers are local and internal of this monolithic system and are therefore intrinsic identifiers.

3.6. EOSC FDO metadata model

Research infrastructures often define their metadata specifications to cater for access and reuse of FAIR (Findable, Accessible, Interoperable, Reusable) data and services. However, due to the multi-disciplinary nature of research, several domain-specific metadata schemas and vocabularies have been established and used. This hinders data and service reuse and discourages and obstructs further and new research.

EOSC aims to enable EOSC users and interoperating services to uniformly find and access data, based on several common metadata attributes. One of the main objectives is for these attributes is to represent datasets drawn from a diverse set of scientific disciplines. To this end, EOSC defines a set of human-understandable, minimal metadata attributes for describing FAIR digital objects, using as a basis predominant metadata scheme, such as the RDA Metadata Interest Group, EOSC pilot, Dublin Core and DataCite. These metadata attributes are illustrated in Figure 2 categorized by their requirement.

The different requirement levels allow for some properties to always be present in the data, such as in the case of mandatory or mandatory if applicable, thus enabling resource discovery in a uniform manner. Recommended or complementary information can be provided using the remaining two levels. Besides uniform resource access, the proposed metadata set enables other software and services to interact with EOSC resources (i.e., resources are machine-understandable). Furthermore, it supports attributes (e.g., project) that aim to capture the management side of the resource. As such, the EOSC minimum metadata set is of prime importance to SLICES-RI and has been utilized in the development of its metadata profiles (see deliverable D4.3). The details of the elements of the EOSC minimum metadata set to describe metadata records for FDO objects can be found in EOSC Interoperability Framework report.

3.7. EOSC metadata Profiles

The EOSC Profiles are part of the EOSC Portal Interoperability Framework that also comprises of the EOSC Portal APIs, the EOSC Portal Onboarding Process (OP), the EOSC Rules of Participation (RoP), together aiming to improve interoperability and promote uptake of research resources that European researchers could explore across Europe. The current version includes two profiles, the Provider and Resource profiles, each addressing a different entity and a different phase of the onboarding, update, maintenance, and monitoring processes of a resource by a provider. Both profiles include specific sets of attributes accompanied by their definitions, participation constraints and validation rules.

---

[19] ORCID: Connectig research and researchers, https://orcid.org/, [Last accessed 27 August 2021]

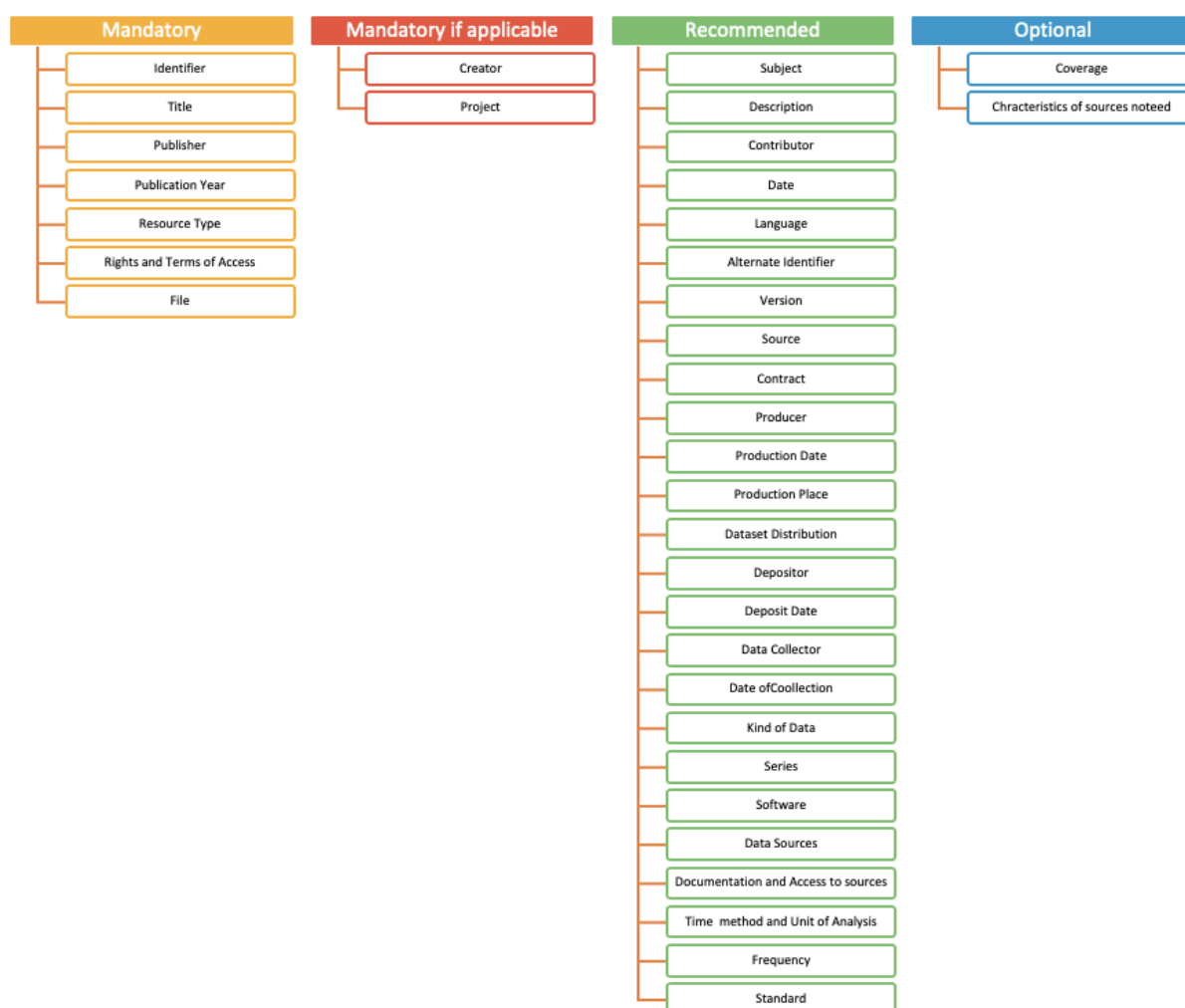*Figure 2 – EOSC Minimal Metadata Attributes and their Requirement Levels*

Figures 3 and 4 provide an overview of the provider and resource profiles' categories and attributes respectively.

Additionally, EOSC provides a structured data entry process using provider and resource code lists, appropriate taxonomies and classifications, which also constitute the basis for the structure and the filtering functions of an EOSC Portal.

Basic information
- ID, Name, Abbreviation, Website, Legal Entity, Legal Status

Marketing information
- Description, Logo, Multimedia

Classification information
- Scientific Domain, Scientific Subdomain, Tags

Location information
- Street Name and Number, Postal Code, City, Region, Country

Contact information
- Main Contact Provider (First Name, Email, Phone, Position), Public Contact (First Name, Email, Phone, Position)

Maturity information
- Life Cycle Status, Certifications

Other information
- Hosting Legal Entity, Participating Countries, Affiliations, Networks, Structure Type, ESFRI Domain, ESFRI Type, MERIL Scientific Domain, MERIL Scientific Subdomain, Areas of Activity, Societal Grand Challenges, National Roadmaps

Figure 3 – EOSC Provider Profile

Basic information
- ID, Name, Resource Organisation, Resource Providers, Webpage

Marketing information
- Description, Tagline, Logo, Multimedia, Use Cases

Classification information
- Scientific Domain, Scientific Subdomain, Category, Subcategory, Target Users, Access Type, Access Mode, Tags

Geographical and Language Availability Information
- Geographical Availability, Language Availability

Resource Location Information
- Resource Geographic Location

Contact Information
- Main Contact Provider (First Name, Email, Phone, Position), Public Contact (First Name, Email, Phone, Position), Other Contact (Helpdesk Email, Security Contact Email)

Maturity Information
- Technology Readiness Level, Life Cycle Status, Certifications, Standards, Open Source Technologies, Version, Last Update, Change Log

Dependencies Information
- Required Resources, Related Resources, Related Platforms

Attribution Information
- Funding Body, Funding Program, Grant/Project Name

Management Information
- Helpdesk Page, User Manual, Terms Of Use, Privacy Policy, Access Policy, Service Level, Training Information, Status Monitoring, Maintenance

Access and Order Information
- Order Type, Order

Financial Information
- Payment Model, Pricing

Figure 4 – EOSC Resource Profile

The EOSC Provider is an EOSC System User (human or machine) responsible for the provisioning of one or more EOSC Resources to the EOSC or exploiting the EOSC System according to the EOSC Policy. EOSC Resources include services, datasets, software, support, training, consultancy, or any other asset. EOSC Providers are organisations, a part of an organisation or a federation that manages and delivers Resources to End-Users. EOSC Providers can be: Resource Providers, Service Providers, Data (Source) Providers, Service Developers, Research Infrastructures, Distributed Research Infrastructures, Resource Aggregators, Thematic Clouds, Regional Clouds, etc. SLICES-RI will be a multi-

modal provider to EOSC, besides the research infrastructure provider role and the ability to execute complex experiments. It will support other provider roles, such as data source provider, service provider or software provider, exposing the data and services to the EOSC research community.

3.8. Conclusion of EOSC IF

Here, we summarize the key features of the EOSC interoperability framework as the following:

(i) EOSC interoperability framework is a set of *not-so-specific* guidelines to ensure smooth integration of infrastructure services and seamless exchange of research data across the EOSC ecosystem. EOSC-IF is derived from the European Interoperability framework which defines interoperability of an information technology system by four key elements, i.e.., technical, semantic, organization and legal interoperability;

(ii) The FAIR principles, federated resource and user management and legal compliances pertaining to privacy, licensing and governance by European Commission are at the core of the EOSC-IF framework;

(iii) In terms of implementation, EOSC FDO is defined as the basic building block for describing EOSC services and data, which should embed all four layers of interoperability;

(iv) EOSC does not (re)invent any new mechanisms and techniques to ensure technical (service specifications, AAI, PID, etc) and semantic (semantic artifacts, metadata models) interoperability, but rather, it provides a framework that advocates to utilize the open specifications for implementation of various building blocks to facilitates technical and semantic level interoperability. For example, EOSC intends to use the popular AARC-BPA for AAI, provides a minimum metadata model that is derived from the existing metadata models such as Dublin core and recommends CrossRef XML schema for the storage and distribution of scholarly publications.

From SLICES point of view, EOSC-IF provides guiding principles that would be highly useful in defining the SLICES interoperability framework to ensure its integration with EOSC for publication of SLICES services and enable the research data exchange.

4. SLICES interoperability framework and integration with EOSC

4.1. Requirements of SLICES-IF

Since SLICES aims to provide a distributed pan-European experimental research platform by jointly utilizing the geographically dispersed computing, storage and networking RIs, it is highly important that the different nodes interacting in the experimental workflow are interoperable with each other. For example, considering a mobile edge computing use case, compute, storage and networking resources from different nodes would be used. In such a scenario, it is necessary that resource description, availability, execution and data exchanges are smooth. This can only be assured if a common interoperability framework is adopted across SLICES so that different subsystems have a common understanding of resources, data/metadata and are on the same page with respect to the licensing, copyright and privacy requirements.

The SLICES-IF mainly targets four communities:

- **Users(research) community**: The SLICES-RI would provide a rich set of testbeds (e.g., 5G/6G, IoTs, etc.), infrastructure services and research data that can be accessed by research communities and industries across the European research area. For researchers accessing SLICES, it is important

that a smooth navigation across different SLICES resources and integration of resources is possible to accomplish the research/experimental workflow. Further, it is important to ensure that the research data (produced as a result of experimentation by the SLICES user or existing data from other SLICES users) conforms to the FAIR principles;

- **EOSC:** EOSC being the flagbearer of future European open digital sciences is a top priority for SLICES due to: (i) publishing of SLICES RIs, services and data through the EOSC portal for a wider reach; and (ii) SLICES users will benefit from the rich set of EOSC services that are expected to even grow more in the future;

- **External RIs:** It is pragmatic to think that some large-scale complex experiments may require SLICES resources to be clubbed with the external RIs such as the public clouds. This means the SLICES-IF should ensure that the state-of-the-art and widely accepted approaches for technical and semantic layers implementation are adopted to ensure a wider interoperability.

- **SLICES consortium partners:** Members of SLICES consortium will provide different resources (hardware and software services) pertaining to computing, data, networking, storage and domain specific resources/testbeds. The interaction of these resources are needed to execute complex scientific experiments. Therefore, a common framework for interoperable resource description, communication protocols, resource integration, semantics and data/metadata models are required;

To achieve this, likewise the EOSC-IF, the SLICES-IF will be built upon the foundations led by the European Interoperability Reference Architecture (EIRA), where interoperability is classified at four layers, namely: (i) technical, (ii) semantic, (iii) organizational; and (iv) legal. Although the target audience for EIRA (governance and administration) was very different of what compared to SLICES, core principles and objectives remain the same. Additionally, the different components (in particular technical and semantic) of SLICES-IF would be chosen in a such way that SLICES is fully interoperable with EOSC for uninterrupted data exchange pertaining to the use of EOSC services and research data by SLICES as well as to enable the publications of SLICES infrastructure, services and data through EOSC portal. Considering EOSC, the following types of data would be exchanged *to-and-fro* between SLICES and EOSC:

- **User data:** For any interaction between SLICES users and EOSC, authentication and authorization of SLICES users is a preliminary step. This would require either exchange of users' data or security tokens/certificates depending on the AAI schemes and cooperation mechanism between EOSC and SLICES infrastructures;

- **EOSC services data:** EOSC provides a plethora of digital tools and services targeting various scientific domains ranging from data processing to biological science and environmental sciences to astronomy. These tools can be highly useful for many future SLICES users. Therefore, it is important that EOSC services can be smoothly integrated into the SLICES workflow to reap the benefits of EOSC services;

- **SLICES resources data**: EOSC being the leading initiative on European digital sciences, SLICES aims to publish its services through EOSC portals due to its wider reach across research communities. To ensure this, SLICES service specifications should conform to the service specifications implemented by EOSC;

- **EOSC Research data:** EOSC ecosystem is expected to produce a huge amount of research data from varieties of scientific domains. This research data can be utilized by SLICES users for several purposes including conducting new experiments, reproducing the existing scientific experiments and for algorithmic benchmarking, etc. In order to consume the research data produced by the EOSC experiments and allow a smooth integration with the SLICES experimental workflow, SLICES need to adopt data models that are compatible with the EOSC data models;

- **Research data and publications produced by SLICES users:** Research publications are important outcomes of scientific experiments. In many cases, research publications are accompanied by models and datasets forming research artifacts that are quintessential for reproducibility of scientific experiments. The large-scale scientific experiments performed through SLICES will generate a wealth of research data/models that can be further used by other academic and industry researchers/ scientists. To maximize the reusability and extensibility of this data, it is important that the data should be stored in formats that can be easily accessed and processed by researchers and developers;
- **Metadata:** Metadata is key to ensure the discovery/findability of scientific resources and research data. Also, metadata contains vital contextual information about the scientific experiments such as the environment in which the experiment was conducted such as the information of the hardware used, software environment and dependencies, etc.

Based on the above discussion, the key requirements of SLICES-IF can be summarized as the following:

1. SLICES-IF is required to be fully compatible with EOSC-IF for user and research data exchange. This requires federated AAI and well defined semantic and metadata catalogues adhering to EOSC specifications;
2. It is important that SLICES adopt open specifications for the representation of services and tools. This is necessary to ensure that tools and services available under the EOSC ecosystem are easy to integrate within the SLICES experimental workflow;
3. A common PID framework for SLICES resources (infrastructure, tools, models, catalogues and data) is required;
4. Easy to use APIs for access through the web as well as command line interfaces are required for user and resource management;
5. The organizational framework should ensure clear rule of governance and participation;
6. The legal framework must provide unambiguous riles for licensing, data use, copyrights and privacy.

4.2. SLICES-Interoperability Framework

Similar to EOSC, FAIR digital object would constitute the basic building block of SLICE-IF with provisions for interoperability at all the layers. Figure 5 below shows the SLICES-IF in the context of EOSC-IF and FAIR digital object guidelines. To achieve interoperability at each layer of technical, semantic, organizational and legal, corresponding enablers fulfilling the concerned requirements will be adopted. For example, to achieve semantic interoperability, SLICES-IF would clearly define the semantic artifacts and metadata following the FAIR principles. A clear hierarchy for semantic artifacts will be developed starting from the less formal representation such as XML schemas and UML models to the highly formal representations such as ontologies. For example, an object class represented in a UML model (less formal) can be linked to an ontology (more formal) which defines the concept of that particular class. Further, SLICES will adopt a generic (domain-agnostic) metadata framework such as GSIM (Generic statistical model) or ISO11179 with provision for domain-specific extensions. The reference(generic) metadata framework will allow semantic representation at the different granularities and can be evaluated to determine the interoperability at different levels. For example, ISO11179 defines two levels to define a variable that are called *data element* and *data element concept* where mapping between these two enables semantic mapping. For example, *data element concept* (e.g., processor) can be related to two *data elements* (e.g., GPU and CPU).
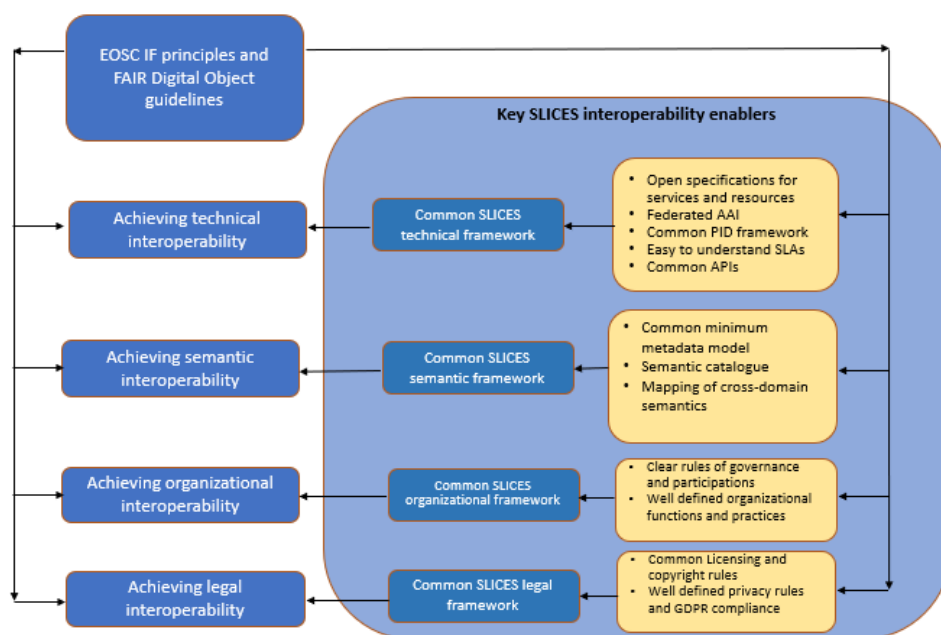
Figure 5 – SLICES Interoperability framework and its interaction with EOSC-IF

Figure 6 shows the conceptual view of SLICES interoperability architecture with different components of SLICES-IF that support the different services (core and peripheral) envisaged. The presented components of SLICES-IF are generic in nature and the specific implementation choices will be decided in the future. It will be ensured that the specific implementation choices will conform to the EOSC-IF. In the following, we describe how interoperability in terms of four interoperability layers will be enabled.

Semantic Design: SLICES will ensure semantic interoperability by developing clear definitions of metadata models and semantic artifacts fully aligned with the FAIR digital object guidelines. Since SLICES aims to support complex scientific experiments from multiple scientific domains, the smooth exchange of information between different domain-specific metadata/semantic models will be enabled by the development of translation/mapping modules of different metadata standards and semantic artifacts.

Technical Design: From the technical point of view, SLICES-IF will be realized using state-of-the-art frameworks and tools of federated AAI management, PID allocation, data access and experiment workflow management. The technical design will ensure that the open specifications for services and infrastructures representation will be used and made compatible with the EOSC-IF implementations and open science initiative. Extensibility will be the key feature of SLICES technical implementation to make it flexible/compatible with the evolution of the technical components and EOSC digital infrastructure.

Legal and organizational design: As explained earlier, this document is primarily focused on technical and semantic interoperability. A detailed description of the organizational and legal interoperability framework is presented in the deliverables D3.1 in the section about the Data Governance Framework.
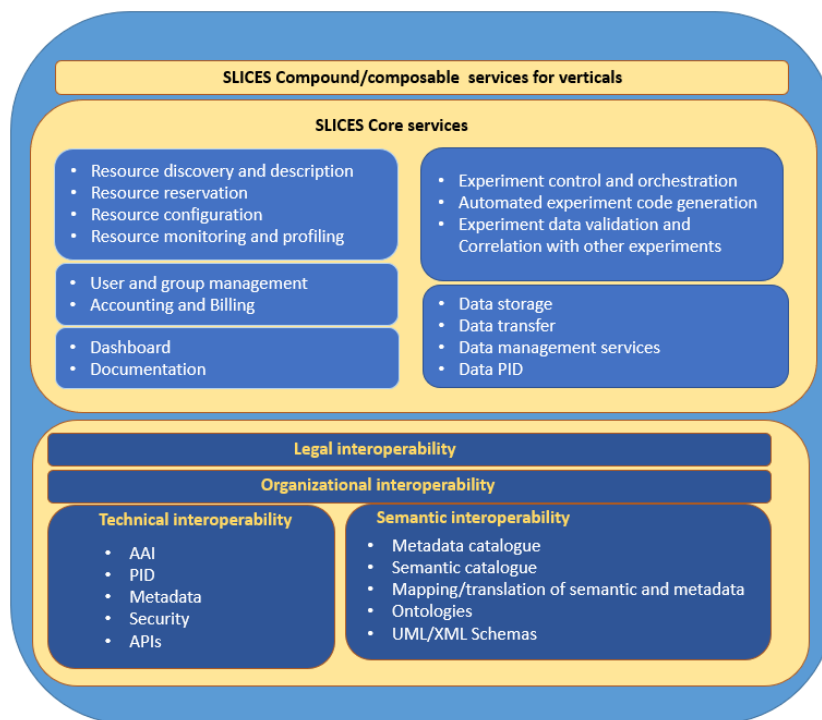
Figure 6 - Conceptual view of SLICES Interoperability Architecture

5. SLICES interoperability components/Design recommendations

5.1. SLICES AAI

There are several standards and open technologies that provide an Authentification and Authorization Infrastructure (AAI). For example, SAML 2.0, OAuth 2.0, OpenID Connect and X.509v3 facilitate interoperability and integration with the existing AAIs of e-Infrastructures, Research Infrastructures, and research communities.

SAML 2.0 is an OASIS standard. SAML 2.0 is an XML-based protocol that uses security tokens containing assertions to pass information about a principal (usually an end user) between a SAML authority, named an Identity Provider, and a SAML consumer named a Service Provider. SAML 2.0 enables web-based, cross-domain single sign-on (SSO), which helps reduce the administrative overhead of distributing multiple authentication tokens to the user.

OAuth 2.0 (Open Authentication) is an open standard for access delegation developed by the IETF OAuth WG. OAuth provides clients a "secure delegated access" to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without providing credentials. Designed specifically to work with Hypertext Transfer Protocol (HTTP), OAuth essentially allows access tokens to be issued to third-party clients by an authorization server with the approval of the resource owner. The third party then uses the access token to access the protected resources hosted by the resource server.

OpenID Connect is an open standard and decentralized authentication protocol, in particular promoted by the OpenID Foundation. It allows users to be authenticated by cooperating sites (known as relying parties, or RP) using a third-party service, eliminating the need for administrators to provide

their own ad hoc login systems, and allowing users to log into multiple unrelated websites without having to have a separate identity and password for each. Users create accounts by selecting an OpenID identity provider and then use those accounts to sign onto any website that accepts OpenID as authentic users. OpenID Connect can be seen as an identity layer on top of the OAuth 2.0 protocol. In OpenID Connect, OAuth 2.0 capabilities are integrated with the protocol itself.

X.509v3 is a standard defining the format of public key certificates. It is defined by the International Telecommunications Union's "Standardization Sector" (ITU-T), in ITU-T Study Group 17 and is based on ASN.1, another ITU-T standard. X.509 certificates are used in many Internet protocols, including TLS/SS. An X.509 certificate contains a public key and an identity (a hostname, or an organization, or an individual) and is either signed by a certificate authority or self-signed. When a certificate is signed by a trusted certificate authority or validated by other means, someone holding that certificate can rely on the public key it contains to establish secure communications with another party, or validate documents digitally signed by the corresponding private key.

Some EOSC-related projects are testing AAI technologies. As an example, the EOSC-hub Authentication and Authorisation Infrastructure (AAI) enables seamless, authenticated access to services and research data. It enables service providers to control access to their services from users holding identities (usernames and passwords) from a very broad set of academic, community or social Identity Providers (IdPs). The EOSC-hub AAI brings together these IdPs, the EOSC-hub service providers (SPs) and intermediary identity management proxies into a single, interoperable infrastructure. SAML 2.0, OpenID Connect, OAuth 2.0 and X.509v3 are in particular supported within EOSC-hub AAI.

In 2021, the EOSC AAI task force publishes a report dealing with the EOSC Authentication and Authorization Infrastructure: EOSC AAI First Principles & Requirements, EOSC AAI Baseline Architecture, EOSC AAI Federation participation guidelines (participation policy and technical framework), and EOSC AAI Best Practises. In particular, it defines interoperability requirements for SAML 2.0, and OpenID Connect/OAuth.
The EOSC AAI Task Force continues its work to deliver a consistent architecture for authentication, authorization and access control for the EOSC following and building upon SRIA and the previous EOSC AAI TF outcomes as defined in this charter.

A remaining question is the choice of Identity Providers and how much they can be trusted. SLICES policies may depend on a confidence level to grant authorization accesses, i.e., to deal with several levels of data confidentiality.

5.2. SLICES FDO and PID Interoperability

Consistent implementation of the services and metadata supporting FDO and PID is an important aspect of the SLICES interoperability and integration with EOSC and other infrastructures. The following aspects should be considered and further investigated towards the final SLICES Interoperability Framework definition and future implementation in the SLICES-RI:

- PID services and roles (refer to main services roles mentioned in section 3.3);

- FDO supporting services such and metadata and ontology registries;

- PID and metadata registries and resolution services;

It can be assumed that at the full implementation and operational stage, SLICES-RI should have all FDO and PID supporting services fully interoperable and integrated with the corresponding EOSC services

and other RI services. However, it is important to identify a minimum viable service profile (MVP) that would allow effectively share and exchange data with EOSC and other EOSC integrated RIs.

The following should be addressed at the design stage and the first implementation phase that should provide full interoperability with EOSC:

- Definition of the metadata profile, including metadata schema and semantics;

- Definition of the PID schema including PID type, PID schema and syntax;

- API to support PID and FDO query and access to digital object or data;

- PID and FDO management services and tools that should support the main operations on FDO: creation, publishing, access/read, update, delete, and correspondingly allow PID creation, publishing and resolution.

Given above, the SLICES MVP may implement the following PDI roles (that can be implemented via corresponding APIs and policies):

- FDO/PID owner;

- PID user/consumer;

- PID Manager;

Other roles such as PID manager, PID local or global provider (such as sectoral PID segment), PID global resolver can be implemented the later stage following the growth of SLICES-RI and volume and variety of data produced.

As an additional interoperability aspect, PID policy governance, compliance and management, in particular interoperability and integration with EOSC PID services and infrastructure, must be reflected in the SLICES Governance framework.

5.3. SLICES metadata profiles

Several metadata specifications, such as Dublin Core, Datacite, DDIand ISO 19115, have been created to address the needs of specific research domains. Some of these standards have been most often associated with a particular file format specification used by a particular community. These standards typically include metadata elements or attributes, accompanied by their definitions and appropriate vocabularies/controlled lists for input validation. Usually, each community maintains the metadata standard and associated rules and procedures for access and discovery as well as the vocabularies used. However, problems of interoperability with different metadata standards arise when a community or system attempt to share and utilize content with external communities or systems. Well-structured metadata and appropriate crosswalks between different metadata structures play a fundamental role in achieving effective interoperability.

Interoperability problems are amplified by big data constraints and the growing demand for cross-disciplinary research that require efficient machine-actionable interoperability, which can offer seamless and precise retrieval of digital objects. It is for that reason that initiatives, such as Data Catalog Vocabulary (DCAT), aim to facilitate interoperability between metadata standards and vocabularies to enable data integration and analysis. The FAIR movement is also aligned towards that goal and has given rise to new initiatives, projects and tools, such as the FAIRsFAIR project and the FAIR Digital Objects Forum which aim to achieve an ecosystem of FAIR infrastructures and systems.

SLICES wants to fully endorse and adopt the FAIR principles through appropriate metadata profiles that enable efficient and effective interoperability and cross-disciplinary research. Due to the problems rising from the multi/cross/inter-disciplinary nature of research SLICES is aiming to support, opting for a union of several established, domain-specific metadata schemas and vocabularies is impractical and would decrease the efficiency and effectiveness of resource discovery, access and (re)use. Instead, SLICES aims to allow its users (and interoperating platforms) to uniformly find, and access any object, such as data, services and software.
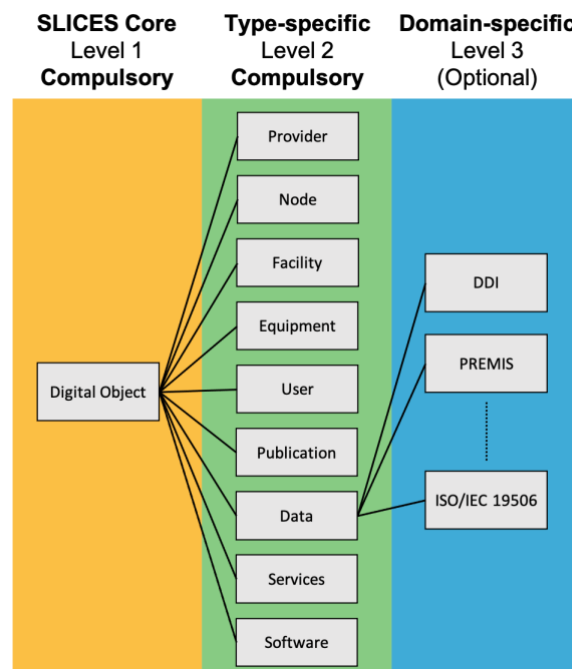

Figure 7 – SLICES Metadata Structure

To accomplish this, SLICES proposes a hierarchical metadata model consisting of three levels, as illustrated in Figure 1. Level 1 includes metadata related to the **S**LICES core **F**AIR **D**igital**O**bject, coined S-FDO. It includes basic information, such as identification, description and its resource type. Management information is also included such as version and metadata profile used. Using its type (e.g., data, service), Level 2 provides type-specific information. For example, a dataset may have start and end date, a facility may have an address. Finally, level 3, provides further domain-specific information that may be required by a specific community. For example, specific media details can be provided to as part of the PREMIS metadata format.

5.4. Proposal for SLICES APIs

Regarding the configuration of the hardware resources, high-level APIs will be provided for exposing the infrastructure to the central hub or external resources, allowing their interconnection and advertisement of resources. As an example, the following two APIs illustrate the need for configuring the experiments: the ETSI NFV SOL005 and the TMForum 620 for lifecycle management of catalogue entries.

ETSI NFV SOL005
The REST API of the ETSI NFV SOL005 interface is currently used for interacting with an NFV-MANO compatible architecture through REST interfaces. The interface specifies all the RESTful APIs for the

interaction with the architecture for the following components: 1) Network Service Descriptor (NSD) Management interface (as produced by the NFV Orchestrator - NFVO) towards the OSS/BSS), 2) Network Service (NS) Lifecycle Management interface (as produced by the NFVO towards the OSS/BSS), 3) NS Performance Management interface (as produced by the NFVO towards the OSS/BSS), 4) NS Fault Management interface (as produced by the NFVO towards the OSS/BSS), 5) VNF Package Management interface (as produced by the NFVO towards the OSS/BSS). The design of the protocol and data model for the above interfaces is based on the information model and requirements defined in ETSI GS NFV-IFA 013[20]. Various similar solutions are emerging and will be explored.

TMForum-620 – Catalog Management

The REST API standardized by TM Forum provides the respective interfaces for managing the entire lifecycle of the catalogue elements, and the consultation of catalogue elements during several processes. The interface can be used for managing the lifecycle of the respective experiments per each experimental island. APIs exist for the configuration of the resource elements in the respective catalogues of each testbed, the catalogue of resources, importing/exporting them, and further interaction with them. Such APIs can be used from the central SLICES hub in order to manage the respective experiments that can run/be on-boarded on each testbed.

6. Adherence to the Open Science principles

Open Science is one of the central aspects and requirements in the new Horizon Europe programme 2021-2027. Open Science is defined as an approach to the scientific process based on open cooperative work, tools and diffusing knowledge. The concepts of Open Science, Open Innovation, Open to the World are intended to ensure excellence and impact of the European Union´s investment in research and innovation, while safeguarding the Union´s interests (Recital 7 Horizon Europe Regulation[21].

Open Science practices include the following activities:

- Providing open access to research outputs (e.g., publications, data, software, models, algorithms, and workflows) through deposition in trusted repositories;

- Publications are expected to be open access; datasets expected to be FAIR and 'as open as possible, as closed as necessary';

- Early and open sharing of research (for example, through preregistration, registered reports, pre-prints, or crowd-sourcing), with the following options both for publications and research data:

    o At the latest upon publication, deposition of the AAM (Author Accepted Manuscript) or VoR (Version of Records) (or publisher PDF – where acceptable) in a trusted repository supplied with the immediate open access via the repository under CC BY or equivalent (CC BY-NC/CC BY-ND are allowed for long-text formats);

    o Information via the repository about any research output/tools/instruments needed to validate the conclusions of the scientific publication;

    o Beneficiaries (or authors) must retain sufficient intellectual property rights to comply with the OA requirements;

---

[20] ETSI GS NFV-IFA 013: "Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Os-Ma-Nfvo reference point - Interface and Information Model Specification".

[21] Establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013, Regulation (EU) 2021/695 of the European Parliament and of the Council, of 28 April 2021 [online] https://eur-lex.europa.eu/eli/reg/2021/695/oj, [Last accessed 27 August 2021]

- o Metadata must be open under CC 0 or equivalent, in line with the FAIR principles and provide information about the licensing terms and persistent identifiers, amongst others.

- Research output management including research data management. Early defined and implemented Data Management Plan (DMP) and its update through the project development are central are element of Open Access to research data;

- Measures to ensure reproducibility of research outputs;

- Participation in open peer-review;

- Involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science).

Additional requirements to research data management in the Horizon Europe projects:

- Beneficiaries must provide (digital or physical) access to data or other results needed for validation of the conclusions of scientific publications, provided legitimate interests safeguarded and unless (open) access already provided at publications with linked data;

- Special case constitutes a public emergency (like in case of COVID19 pandemic data, or other emergency situations), beneficiaries must (if requested by the granting authority) immediately deposit any research output in a repository + provide open access to it under CC BY, CC 0 or equivalent;

- As an exception, if the access would be against the beneficiaries' legitimate interests, beneficiaries must grant non-exclusive licenses — under fair and reasonable conditions — to legal entities that need the research output to address the public emergency and commit to rapidly and broadly exploit the resulting products and services at fair and reasonable conditions;

- This provision applies up to four years after the end of the action.

Until recently, there were two main services to support Open Science for European research OpenAIRE[22] and Zenodo[23] that are highly regarded by European researchers and academia.
To effectively support Open Science, a new service Open Research Europe (ORE - https://open-research-europe.ec.europa.eu/) has been launched by EC in March 2021 for H2020 and HE beneficiaries. ORE ensures high scientific standards (e.g., editorial policies and guidelines), expert Scientific Advisory Board across all fields of science. Benefits include swift publication times and transparent processes (e.g., open peer-review) with no cost to authors/beneficiaries (publication fees paid by the Commission). ORE is a recommended venue for publishing results of the European projects/grants, however no obligation to publish there but benefit is automatic compliance with HE policy.

> **Recommendation:** Adherence to the Open Science principles will naturally happen in SLICES as this requirement is non-separable requirement in Horizon Europe grants and this requirement is facilitated by requirements in the project proposal templates, reviewing process and assessment process. Advance planning of Open Science compliance should take place at the design stage and be embedded into the SLICES Governance and Management structure.

---

[22] OpenAIRE, https://www.openaire.eu/, [Last accessed 27 August 2021]
[23] Zenodo platform, https://zenodo.org, [Last accessed 27 August 2021]

7. Conclusion

This document presents a vision for the SLICES interoperability framework to facilitate the integration of SLICES with the EOSC. The main purpose of SLICES-IF is to allow the publication of SLICES services on EOSC portal, enable a smooth exchange of research data and services between EOSC and SLICES and serve the federation of RIs that constitute SLICES by providing common frameworks/models of AAI, PID policy, APIs, semantic artifacts and metadata representation. As the starting point, a detailed study of EOSC-IF was carried out to understand the technical and semantic aspects of the interoperable exchange of data and services within the EOSC ecosystem. Besides, the proposed use cases of SLICES-RI were analysed in the context of SLICES as a distributed RI with geographically distributed nodes to derive the key interoperability requirements for carrying the complex scientific experiments. Based on the recommendations of EOSC-IF and the requirements of the SLICES use cases, SLICES-IF is defined. Similar to the EOSC-IF, FAIR digital object (FDO) is chosen as the basic building block of SLICES-IF where an FDO fully satisfies the interoperability requirements pertaining to the technical, semantic, organization and legal aspects.