Research Lifecycle Management technologies for Earth Science Communities and Copernicus users in EOSC

Deliverable D6.1
EOSC Integration Plan

Grant agreement number	101017501
Start date of the project	01/01/2021
Duration of the project	24 months
Type of Action	Research and Innovation Action
Coordinator	PSNC

Due date of delivery	31/03/2021
Actual date of delivery	31/03/2021
Work package	WP6
Type of deliverable	Report
Dissemination level	Public
Responsible	PSNC
Reviewer	MEEO
Version	1.0

This project has received funding from the European research infrastructures (including e-Infrastructures) *under the European Union's Horizon 2020 research and innovation programme* under grant agreement No 101017501

List of authors, contributors and reviewers

Name	Role	Organization
Raul Palma	Author	PSNC
Soumya Braham	Author	PSNC
José Manuel Gómez Pérez and Andrés García	Contributors	ESI
Simone Mantovani	Contributor	MEEO
Pedro Gonçalves	Contributor	TERRADUE

History of changes

Version	Date	Change	Authors	Organization
0.1	28/02/21	ToC	Raul Palma	PSNC
0.2	10/03/21	Initial input	Raul Palma	PSNC
0.3	15/03/21	Input	Soumya Brahma	PSNC
0.4	20/03/21	Inputs	Raul Palma/ Soumya Brahma	PSNC
0.5	25/03/21	Input	Andrés García and JM Gómez Pérez	ESI
0.6	25/03/21	Input	Simone Mantovani	MEEO
0.7	25/03/21	inputs	Pedro Gonçalves	TERRADUE
0.8	26/03/21	inputs	Raul Palma/ Soumya Brahma	PSNC
0.9	30/03/21	Review	Pedro Gonçalves	TERRADUE
0.95	30/03/21	Review	Simone Mantovani	MEEO
1.0	31/03/21	Updates & formatting	Raul Palma/ Soumya Brahma	PSNC

Glossary

Acronym	Explanation
AAA	Authentication, Authorization and Accounting
AAI	Authentication and Authorization Infrastructure
ADAM	Advanced geospatial Data Management
API	Application Program Interface
C-SCALE	Copernicus - eoSC AnaLytics Engine
CA	Consortium Agreement
CAPM	Capacity management
CHM	Change management
CMF	Cloud Management Framework
CNR	Consiglio Nazionale Delle Ricerche
CONFM	Configuration management
CRM	Customer relationship management
CSI	Continual service improvement management
CSW	Catalogue Service for the Web
DICE	Data Infrastructure Capacity for EOSC
DM	Data Manager
DOI	Digital Object Identifier
EAB	External Advisory Board
EC	European Council
ECMWF	European Centre for Medium-Range Weather Forecasts
EGI	European Grid Infrastructure
EGI-ACE	EGI Advanced Computing for EOSC
EO	Earth Observation
EOSC	European Open Science Cloud
ESI	Expert Systems Iberia
ESWG	EOSC Sustainability working group
EU	European Union
EUMETSAT	European Organization for the Exploitation of Meteorological Satellites
EVEREST	European Virtual Environment for Research Earth Science Themes
FAIR	Findable, Accessible, Interoperable, Reproducible
GA	General Assembly
HTTP	Hypertext Transfer Protocol
ICT	Information and communications technology
IdP	Identity Provider
IM	Innovation Manager
INGV	Istituto Nazionale di Geofisica e Vulcanologia
IP	Intellectual Property
ISM	Information security management
ISMAR	Istituto di Scienze Marine
ISRM	Incident and service request management
ITSM	IT service management system
MEEO	Meteorological and Environmental Earth Observation

MS	Milestone
MSFD	Marine Strategy Framework Directive
MSP	Maritime Spatial Planning Directive
MVE	Minimum Viable EOSC
NEANIAS	Novel EOSC Services for Emerging Atmosphere, Underwater & Space Challenges
NEON	National Ecological Observatory Network
NL	Natural Language
OEG	Ontology Engineering Group
OGC	Open Geospatial Consortium
ORCID	Open Researcher and Contributor ID
PaaS	Platform-as-a-Service
PC	Project Coordinator
PDF	Portable Document Format
PM	Problem management
PMT	Project management Team
PSC	Project Steering Committee
PSNC	Poznan Supercomputing and Networking Center
RDM	Release and deployment management
RELIANCE	Research Lifecycle Management technologies for Earth Science Communities and Copernicus users in EOSC
REST	Representational state transfer
RI	Research Infrastructure
RO	Research Objects
SaaS	Software as a Service
SACM	Service availability and continuity management
SAML	Security Assertion Markup Language
SC	Scientific Coordinator
SDG	Sustainable Development Goals
SDT	Service Description Template
SEA	Strategic Environmental Assessment Directive
SLM	Service level management
SMS	Service Management System
SPM	Service portfolio management
SRIA	Strategic Research and Innovation Agenda
SRM	Service reporting management
SSO	Single sign-on
SUPPM	Supplier relationship management
T2	Terradue
TC	Technical Coordinator
TL	Task Leaders
TRL	Technology Readiness Level
TTS	Token Translation Services
UiO	Universitetet i Oslo
UPM	Polytechnic University of Madrid (Universidad Politécnica de Madrid)
VM	Virtual Machine
VO	Virtual Organization
VRC	Virtual Research Communities
VRE	Virtual Research Environment

WCS	Web Coverage Service
WFD	Water Framework Directive
WG	Working Group
WMS	Web Map Services
WP	Work package
WPL	Work Package Leaders

Table of Contents

List of Figures

List of Tables

1 Executive Summary

RELIANCE, short for Research Lifecycle Management for Earth Science Communities and Copernicus users in EOSC, aims to realize the vision of FAIR research in EOSC by adopting a holistic research management approach based on three key and complementary technologies: i) Research objects (RO) as the overarching mechanism to manage scientific research activities, which relies upon ROHub platform as the reference service; ii) data cubes as the mechanisms enabling an efficient and scalable Earth Observation data discovery and access, which relies upon the Advanced geospatial Data Management (ADAM) platform as the reference service; iii) text mining and semantic enrichment services allowing to extract machine-readable metadata from RO resources, enabling researchers to discover scientific information at scale and to structure their own research, and which rely on the AI-based platform Cogito as base system. As part of the integration in EOSC, RELIANCE services will leverage and integrate with some of the EOSC core-cross cutting and added value services (as described in this document), playing a complementary role to what is already available and bridging between various EOSC services. RELIANCE will pilot the services in three different Earth Science communities, fostering the use of Copernicus data and demonstrating their efficacy in real-life vertical and multi-disciplinary scenarios, and will launch an Open Call to engage other communities.

This document describes the plan for integration RELIANCE services into the EOSC. The plan is based on the existing guidelines and the current EOSC landscape which is a continuously evolving environment. For this, it takes into account the latest results from different EOSC working groups, including the documents delivered previous month (February 2021) describing the EOSC Architecture and Minimum Viable EOSC (MVE), the EOSC interoperability framework and the on-going work of the recently started projects funded along with RELIANCE as part of the INFRAEOSC-07 programme.

The document introduces the core services that RELIANCE will integrate into EOSC, including the ROHub platform, ADAM platform and Text Mining services, as well as the Copernicus Data Pipelines as an additional added-value service. The document also highlights EOSC services that are planned to be leveraged and the roadmap for implementing the integration plan.

The integration plan and its implementation will be assessed, and revised if needed, in the Annual report on EOSC operation in M12 of the project, and the final results will be presented in the Annual report on EOSC operation in M24 of the project.

2 Introduction

2.1 Scope

This deliverable presents the RELIANCE plan for the integration of its services into the EOSC based on the existing guidelines for service providers and the current EOSC landscape. The document provides an overview of the latest EOSC architecture including the definition of the Minimal Viable EOSC (MVE) comprising essential EOSC services that bring value to users beyond their current use of infrastructures. Additionally, the document describes the process for onboarding and integrating services into EOSC, including an overview of other relevant EOSC services that may be leveraged by RELIANCE.
In particular, RELIANCE services will connect to core - cross-cutting - services including the EOSC AAI, as well as other common and added-value services for researchers and/or service providers. This will include the EOSC infrastructure providing storage, cloud computing resources and virtual machines on demand, which would enable RELIANCE services to scale-out and to support a large number of users and research communities. For instance, these communities typically require access to specific services, usually on the Cloud, and/or to specific tools/applications that can be in a Jupyter Notebook or a command line application with different requirements for scalable storage and computing environments. To run such community-specific tools/applications, researchers may need access to virtual machines that can be easily shared/replicated with collaborators. The operation of RELIANCE services, which aims to support multiple research communities in Earth Science, will require access to a highly scalable and dynamic ICT infrastructure that can host the different cloud services needed by the communities, and which can provide scientists with resources to create their own virtual environments. For this, RELIANCE plans to rely on the resources made available via the European Open Science Cloud Compute Platform, including the Notebook facilities that will be used as the processing and analysis environment for research objects.

RELIANCE will also leverage and reuse other EOSC services for research data management, including services for depositing and sharing (B2DROP, B2SHARE, Zenodo), discovery and reuse (Research Graph), and data management (Argos). RELIANCE will act as a bridge between other EOSC services, e.g., connecting the data used by researchers, including Data Cubes for accessing Copernicus data and other data resources deposited in EOSC services (e.g., B2DROP, B2SHARE), the methods used to process such data (e.g., via EOSC Notebooks), and the final results published via scholarly communication services (e.g., Zenodo). Moreover, as part of the research object evolution, different snapshots/releases can be generated throughout the research lifecycle, which can be shared and published via research data platforms like B2SHARE and Zenodo for its long-term preservation and citation.

The services that RELIANCE will integrate into EOSC include:

- ROHub platform providing Research Object management functionalities
- ADAM platform providing Data Cubes management functionalities
- Text Mining services providing functionalities for information extraction and semantic annotation, content-based retrieval and recommendation, as well as extended analytics services in support of the scientific enterprises
- Copernicus Data Pipelines, an added-value service enabling the systematic execution of Earth Observation (EO) applications to continuously deliver data and information to different users, complying with their specifications for information content and format.

The goal of this deliverable is to specify the integration plan, including a roadmap to implement this plan. The document will serve to RELIANCE partners as a reference guide with the steps necessary to onboard and integrate their services into the EOSC. The roadmap will allow RELIANCE to reach and assess the project milestones.

The deliverable structure and content takes as starting point existing guidelines for service providers, including the 'EOSC-hub Integration Handbook for Service Providers'[1] [1] and other similar recent reports like the NEANIAS deliverable D8.1 EOSC integration plan [2], and it takes into account the latest results from the different EOSC working groups, including the EOSC Architecture Working Group View on the Minimum Viable EOSC (MVE) report [3], the EOSC interoperability framework report [4] and the on-going work of the recently started projects funded along with RELIANCE as part of the INFRAEOSC-07 programme.

2.2 Audience

This deliverable is intended for internal use by the RELIANCE Consortium, although it may be valuable to external stakeholders, including other EOSC related projects who are also dealing with the integration of their services into the EOSC.

2.3 Structure

The rest of the document is structured as follows:

- Section 3 provides an overview of the EOSC and the current landscape around it. This includes an overview of directly related EOSC projects and working groups.
- Section 4 provides an overview of the EOSC architecture including the definition of the Minimum Viable EOSC (MVE), and the list of the associated services in the first phase (2021-2023).
- Section 5 provides an overview of the EOSC guidelines for the onboarding and integration of services in the EOSC. This takes into account the latest documents and handbooks for service providers that are provided by EOSC.
- Section 6 provides an overview of the RELIANCE services that will be onboarded and integrated in the EOSC, including: ROHub platform, ADAM platform, Text Mining services, and the Copernicus EO Data Pipelines.
- Section 7 provides an overview of the EOSC services that are planned to be leveraged by RELIANCE, with a discussion of how they will be used and integrated
- Section 8 provides the roadmap to implement the integration plan, including an overview of the related tasks.
- Section 9 finishes with the conclusions

[1] https://doi.org/10.5281/zenodo.3826907

3 European Open Science Cloud (EOSC) overview

The European Open Science Cloud (EOSC)[2] is a European Commission (EC) initiative aiming to develop a trusted, virtual, federated environment in Europe to store, share, process and re-use research digital objects (publications, data and software) across borders and scientific disciplines, for research, innovation and educational purposes, following FAIR principles[3] [5][4]. The EOSC brings together institutional, national and European stakeholders, initiatives and data infrastructures to develop an inclusive EU open science ecosystem that aggregates services, software, data and other types of scientific outputs from a diverse set of providers.

The EOSC initiative started in 2015[5] as an EC vision of a large infrastructure to support and develop open science and open innovation in Europe and beyond by federating existing EU research data infrastructures and by realizing a web of FAIR data and related services for science that would make research data interoperable and machine actionable. After being endorsed in May 2018 by the EU committee on research, the EOSC officially launched in November 2018, starting to provide access to services via their EOSC Portal[6].

EOSC aims to offer 1.7 million European researchers and 70 million professionals in science, technology, the humanities and social sciences a virtual environment with open and seamless services for research, innovation and educational purposes, as well as to support EU science in its global leading role. Its particular goals include[7]:

- foster best practices of global data findability and accessibility;
- help researchers get their data skills recognized and rewarded;
- help address issues of access and copyright and data subject privacy;
- allow easier replicability of results and limit data wastage;
- contribute to clarification of the funding model for data generation and preservation, reducing rent-seeking and priming the market for innovative research services.

The development of EOSC during the first phase as part of Horizon 2020 (H2020) programme[8] (2018-2020), was governed by three bodies:

- EOSC Governance Board: an institutional group with representatives from EU countries, countries associated with Horizon 2020 and the Commission that ensure effective supervision of the EOSC implementation.
- EOSC Executive Board: a body with representatives from the research and e-infrastructures communities that to ensure implementation and accountability
- EOSC stakeholders: a forum with a wider range of actors, consulted through a series of stakeholder events and online consultations to collect input and recommendations.

Under Horizon Europe[9], the Commission's research and innovation funding programme, succeeding Horizon 2020 from 2021, EOSC is running as a co-programmed European Partnership. In July 2020, an

[2] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en
[3] https://www.force11.org/group/fairgroup/fairprinciples
[4] https://www.nature.com/articles/sdata201618
[5] https://www.egi.eu/about/newsletters/what-is-the-european-open-science-cloud/
[6] https://eosc-portal.eu/
[7] https://www.eoscsecretariat.eu/news-opinion/european-open-science-cloud-what-it-about
[8] https://ec.europa.eu/programmes/horizon2020/en
[9] https://ec.europa.eu/info/horizon-europe_en

EOSC association was set up to provide a single voice for advocacy and represent the broader EOSC stakeholder community. It is currently starting operations and will rapidly expand its membership.

In this second phase (after 2020), the new governance model for the next EOSC implementation phase includes:

- the EU represented by the Commission
- the EU research community represented by the EOSC Association
- EU countries and associated countries (under Horizon Europe) represented through a steering board to be set up in 2021 outside of the EOSC association

The EC is currently providing financial support to implement the EOSC through different H2020 projects. Currently, over 50 H2020 projects are contributing to the EOSC implementation[10].

Initial projects like EOSC-hub[11], EOSCpilot[12], OpenAire[13], eInfraCentral[14] contributed to the establishment of the EOSC central services in 2018, and operate and continuously evolve them since then. EOSC-hub, for example, delivered the integration and management system (the Hub) of the EOSC that delivers a catalogue of services, software and data from the EGI Federation, EUDAT CDI, INDIGO-DataCloud and major research e-infrastructures. The Hub builds on mature processes, policies and tools from the leading European federated e-Infrastructures to cover the whole life-cycle of services, from planning to delivery. The Hub aggregates services from local, regional and national e-Infrastructures in Europe, Africa, Asia, Canada and South America.

These projects started the implementation of the EOSC Portal, which was officially launched in November 2018, and since then contributed with its maintenance and update. Moreover, in December 2019, the EOSC Enhance[15] project started with the mandate of bringing forward the developments of the EOSC Portal. An integral part of the EOSC Portal is the EOSC Marketplace, powered by EOSC-hub, providing a single-entry point to discover, access, use and reuse (order) a broad spectrum of services and resources for advanced data-driven research.

The RELIANCE plan for the integration with the EOSC is based on the current capabilities and possibilities available via the EOSC Portal and its back-end Hub. However, we are also taking already into account the work that is currently being carried out as a part of EOSC related projects and working groups, which include in particular (but not limited to):

- The definition of the EOSC architecture and the Minimum Viable EOSC (MVE) as part of the EOSC Architecture Working Group[16]. The architecture document was delivered just last month (February 2021), and is discussed in the next Section.
- The definition of the EOSC interoperability framework as part of the Interoperability Task Force of the EOSC Executive Board FAIR Working Group[17], with participation from the Architecture WG. The document was also delivered just last month (February 2021), and is taken into account particularly for the integration of RELIANCE services (Section 7).
- The synergies with sister projects that were funded along with RELIANCE under the call INFRA EOSC-07 - Increasing the service offer of the EOSC Portal. These include:

[10] https://eosc-portal.eu/about/eosc-projects
[11] https://eosc-hub.eu/
[12] https://eoscpilot.eu/
[13] https://www.openaire.eu/
[14] https://cordis.europa.eu/project/id/731049
[15] https://www.eosc-portal.eu/enhance
[16] https://eoscsecretariat.eu/eosc-architecture-wg-outputs
[17] https://www.eoscsecretariat.eu/working-groups/fair-working-group

- EGI-ACE[18] (EGI Advanced Computing for EOSC), which will deliver the EOSC Compute Platform and will contribute to the EOSC Data Commons through a federation of cloud compute and storage facilities, PaaS services and data spaces with analytics tools and federated access services.
- DICE[19] (Data Infrastructure Capacity for EOSC), which aims to enable an EU storage and data management infrastructure for EOSC, providing generic services and building blocks to store, find, access and process data in a consistent and persistent way.
- OpenAIRE-Nexus[20] Scholarly Communication Services for EOSC users, which brings in EOSC a set of services to implement and accelerate Open Science, grouped in three portfolios: PUBLISH (catch all repository; Open Access overlay journal platform; data anonymization; Data Management Plans), MONITOR (Open Science and research impact monitoring; open citation indexes for article-article, article-dataset links; EU monitoring of Article Processing Charges, publication usage statistics), and DISCOVER (open catalogue and APIs to the OpenAIRE Research Graph of interlinked publications, data, software, projects; discovery portals for communities; validation and brokering services for data sources to improve their metadata).
- C-SCALE[21] (Copernicus - eoSC AnaLytics Engine), which aims to federate European EO infrastructure services, such as the Copernicus DIAS and others, capitalizing on the EOSC's capacity and capabilities.

- The upcoming EOSC Future project[22] funded under the INFRAEOSC-03 call in order to integrate, consolidate, and connect e-infrastructures, research communities, and initiatives in Open Science to further develop the EOSC Portal, EOSC-Core and EOSCExchange of the European Open Science Cloud (EOSC).

[18] https://www.egi.eu/projects/egi-ace/
[19] https://www.dice-eosc.eu/
[20] https://www.openaire.eu/openaire-nexus-project
[21] http://c-scale.eu/
[22] https://www.isti.cnr.it/en/research/project-detail/13653/EOSC_Future_EOSC_Future

4 EOSC architecture overview

To make the EOSC platform sustainable, the EOSC Sustainability working group (ESWG) published the FAIR lady report[23] [6] in November 2020 in order to explore possibilities to sustain the EOSC beyond the initial phase that terminated at the end of 2020. To showcase the requirement of sustainability, a first iteration of the Minimum Viable EOSC (MVE) was described in that document. Taken this as starting point, the EOSC Architecture Working Group published its view on the MVE[24] [3], addressing lack of details on concrete services.

The EOSC high-level architecture comprises the following terminology:

- **EOSC Core** comprises the set of enabling services required to operate the EOSC platform.
- **EOSC Exchange** is a set of federation services registered to EOSC by the RIs and the clusters in order to serve the needs of the research communities and thus further widened to the public and private sectors.
- **EOSC Federation** is the set of scientific services provided by the RIs and clusters to the respective research communities.
- **Minimum Viable EOSC** is the set of the dynamic EOSC resources targeted to make the EOSC platform sustainable for the widened research communities including:
    - The subset of EOSC resources needed to form the essential added value provided by EOSC, allowing the discovery, composition, access and analysis of essential services and research products via the EOSC.
    - The subset of EOSC-Core components/services required to operate and deliver the above-mentioned resources.

The Figure 1 shows the relationship between the EOSC-Core, EOSC-Exchange, EOSC-Federation and the MVE.

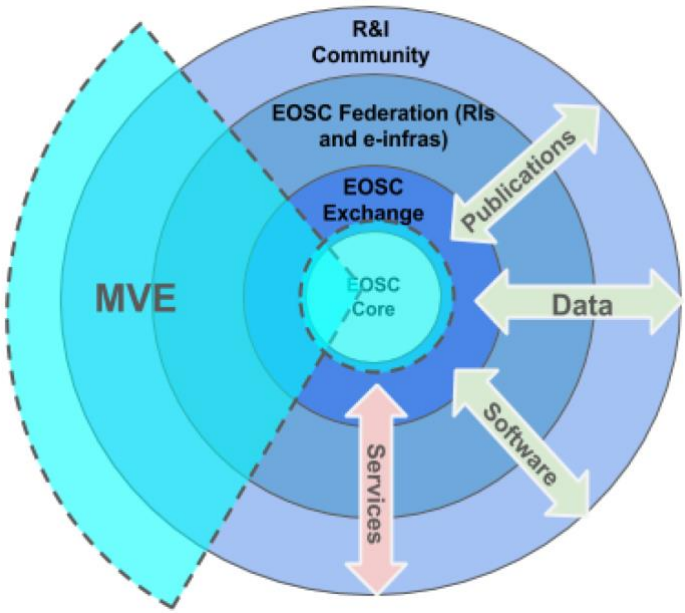

Figure 1: High-level diagram of the EOSC depicting the relationship between EOSC-Core, EOSC-Exchange, EOSC-Federation and the MVE

[23] https://op.europa.eu/en/publication-detail/-/publication/581d82a4-2ed6-11eb-b27b-01aa75ed71a1
[24] https://op.europa.eu/en/publication-detail/-/publication/91fc0324-6b50-11eb-aeb5-01aa75ed71a1

The EAWG described a functional overview of the EOSC and identified frameworks and functions of the EOSC interoperability framework (EIF) to be included in the MVE focusing on essential functions from the EOSC core. The Figure 2 depicts the architecture of the EOSC taking those aspects into account. The diagram provides a functional overview of the EOSC, identifying:

- EOSC Users demand side and EOSC Resource Providers supply side
- EOSC-Core functions and capabilities
- EOSC Interoperability Framework with interoperability guidelines to support the integration and resources across providers and for connecting resources to the EOSC-Core functions.

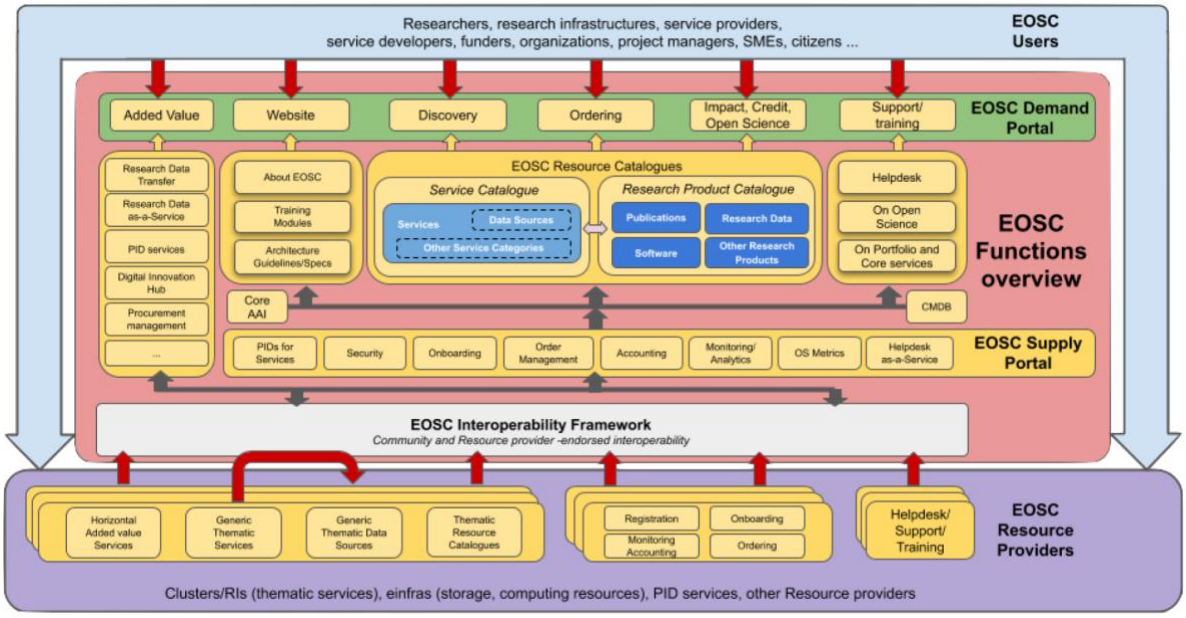

Figure 2: The functional architecture of EOSC

The Architecture WG identified the MVE at the function level, and focuses on the period from now until the end of 2023, corresponding to the phase 1 of the EOSC as defined in the Strategic Research and Innovation Agenda (SRIA)[25]. In the following we list those functions defined as belonging to the EOSC-Core, grouped by Area, and planned in phase 1, constituting the initial MVE.

- PID services
  - PIDs for research entities in EOSC (Part of EOSC core): Policy framework for PIDs for research entities (data, software, publications, organizations, researchers, funders, etc.)
- Portal, Catalogues & Orders
  - EOSC Portal (Part of EOSC core): Website with basic information about EOSC, embeds the catalogue(s) in it.
  - EOSC Portal Service Catalogue and federated portfolio (Part of EOSC core) having a frontend of the portal which exposes services to the customers and a backend to collect the services, publish them and allow onboarding service.
  - EOSC Data source portfolio and onboarding of data sources (registering, validating, discovering) (Part of EOSC core) having frontend that exposed the data sources to customers and a backend to collect the relevant information regarding the data sources and services from the customers.

[25] https://www.eoscsecretariat.eu/sites/default/files/eosc-sria-v09.pdf

- AAI & Security
  - EOSC AAI in support of the EOSC Service portfolio providing infrastructure which allows sharing of login and access to services and data across EOSC.
  - EOSC AAI for EOSC core services providing a SSO environment on the basis of Federated Identities across the EOSC Core services.
  - EOSC Security Policies and Security Coordination Functions are the central coordination of security activities for the EOSC core ensuring compatibility of the identity policies and operational security.
- Helpdesk
  - EOSC portal/core Helpdesk is a basic helpdesk to cover incidents for the portal and for the core services.
- Support Services
  - EOSC Core Collaboration Software is basically a software platform to track documentation, tasks, communication of the operations of the core of EOSC.
  - EOSC Support Services includes consultancy and training on how to use and benefit from EOSC core services
  - EOSC Open Science training and support addressing the technical-organizational and legal aspects of data and service interoperability.
  - EOSC Open Science Help Desk and Collaborative tools aimed for a network of OS experts to be in contact with EOSC users/providers to aggregate European training/support OS material.
- Monitoring, metrics and accounting
  - EOSC Open Science metrics are the infrastructures for scientific communities which gather all types of usage data for all types of resources (citations, usage events for data, services, software) and offer Analytics Services.
  - EOSC Accounting are statistics on services in the EOSC-Exchange to support billing and accounting of services through e.g., Virtual Access.
  - Core Services CMDB includes a database of entities that contribute to delivery of the core services to support change management, monitoring and accounting services.
- EOSC Interoperability Framework
  - Shared open science policy framework for ensuring openness and interoperability, privacy and security.
  - Interoperable metadata framework provides an understanding of metadata schemas, mapping between schemas, registering schemas, ontologies etc.
  - Compliance framework: Rules of Participation, including a Resource Description Framework.
  - Service management and access framework allows services and operational roles in delivering core services and supporting external services that are to be delivered.
  - Compliance Framework: Interoperability policies are a set of policies, recommendations and standards on how EOSC services can be built and interact to have a strong impact on technical services.
  - Compliance Framework: Policies that need to be applied in the EOSC domain to ensure the coherence and value generation of EOSC.

5 EOSC integration guidelines

This section is based on the integration guidelines provided by the EOSC-hub project in the document 'EOSC-hub Integration Handbook for Service Providers' [1][26] and the deliverable D8.1 EOSC integration plan [2] from the NEANIAS project. The section explains the compulsory and the optional steps that service providers must need to qualify as service operators in the EOSC platform.

5.1 Integration overview

The minimum level of EOSC integration is that the service entry should be published in the EOSC Portal and Marketplace. Prior publication the service entry should meet the minimum requirements, it is visible and properly described and accessible for new users. If a particular service requires users to apply for access, then such requests will be submitted via the EOSC marketplace. Next, such requests are forwarded to the service provider who evaluates and responds to the requests using the dedicated EOSC tools. This process is described in detail in subsection 5.2. Besides the publication of the service in the portal, additional integration options are available that are optional, but that can bring further added value to providers and users. Hence, the provider can select from these integration options according to the benefits the users can get from them. These other integration options are described in the later subsections (5.3, 5.4 and 5.5).

5.2 Registering in the EOSC portal (onboarding new service)

The EOSC services can be accessed by users via the EOSC Portal[27]. Any service that providers want to include in the EOSC Service Portfolio needs to follow the onboarding process. Service onboarding is the process whereby a service joins the EOSC service portfolio, and subsequently the list of services in the Marketplace on the EOSC Portal website. Services published in the EOSC portal and marketplace get different benefits, such as promotion of the service to users outside their specific community, a single gateway for users to discover and use services, and the potential integration with other services in the catalogue. Services may also be directly ordered through the EOSC Portal.

The following Figure 3 provides an overview of the on-boarding process.

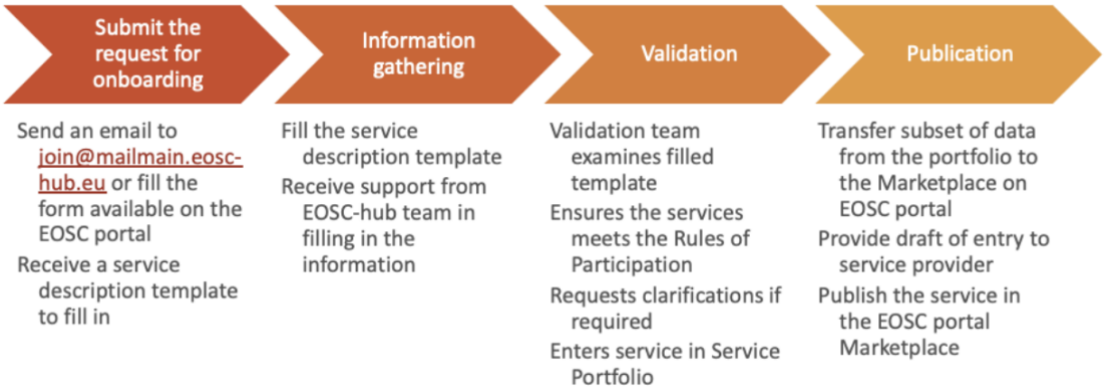

Figure 3: Overview of the onboarding of a new service

[26] https://doi.org/10.5281/zenodo.3826907
[27] https://eosc-portal.eu/

The initial step is to contact EOSC-hub and request for onboarding. This is done via email or by completing the provider form on the EOSC Portal website[28]. This form asks the user to provide the initial information about the service, reason for wanting to become an EOSC service provider, access conditions and applicant contacts. The information submitted becomes a ticket which is used for tracking the request, and a Service Description Template (SDT)[29] to be filled is sent to the provider.

The SDT covers more detailed information such as detailed description of the service, licenses, helpdesk contact, accessibility, etc. for the onboarding team to review and check if the requester's service fulfils all the requirements of the "Rules of Participation" specified for the EOSC portal. Additionally, the information needed to populate the EOSC Service portfolio entry corresponding to the new service is also collected through the SDT. The onboarding team also provides customizable templates for some of the items, e.g., for 'acceptable user policy'. These templates can help providers fill any gaps they are currently missing from a complete service description. The fundamental requirements[30] for a service to be onboarded in EOSC can be summarized as follows:

- The service should fall within the scope of the EOSC activities and brings value in the implementation of Open Science concept.
- It has to be either an online service (e.g., a web application portal, a web service) or a 'human' service, such as training and consultancy.
- In case of a service, it must have at least Technology Readiness Level (TRL)[31] 7
- The compulsory fields of the service description template are to be filled during onboarding.

The submitted information is then reviewed for suitability by the onboarding team and then a draft is set up about the service within the EOSC Portal. This draft is sent back to the submitter for final validation. After successful validation the service entry is made public and is accessible in the EOSC Portal and Marketplace subsequently.

There are also a few additional steps (optional) in case the user wants to make the service orderable through the EOSC portal (currently at TRL 8) where the user can configure the ordering system on the Marketplace platform. Another optional step will be the integration with other EOSC services including the ones described in the following subsections.

5.3 Integration with federation services

This section provides an overview of the set of services offered by the EOSC-hub that helps service providers to enhance their services from the operational perspective. These services can facilitate different operational tasks such as simplifying how users access the federated authentication services, improving the reliability of the services, providing details on resource consumptions, or simplifying user interaction via a helpdesk.

5.3.1 *AAI services*

The EOSC-hub Authentication and Authorization Infrastructure (AAI) enables authenticated access to services and research data in EOSC. The AAI enables service providers like RELIANCE to control access to their own services from users holding identities (usernames and passwords) from a broad set of academic, community or social Identity Providers (IdPs), including those that are part of the eduGAIN

[28] https://eosc-portal.eu/for-providers

[29] https://docs.google.com/spreadsheets/d/1zeUShdnFQU5fTeKSyOcvlCCKeGMA6sbnp7bUkXaa97k/edit#gid=1115091576

[30] https://wiki.eosc-hub.eu/display/EOSC/Criteria+for+possible+inclusion+in+the+EOSC+Service+Portfolio

[31] https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-gtrl_en.pdf

federation, ORCID, Google, Facebook, LinkedIn and others. The AAI brings together the IdPs, EOSC-hub service providers and intermediary identity management proxies into a single, interoperable infrastructure.

The EOSC-hub AAI is built upon open technologies to offer a flexible authentication and access management framework (e.g., SAML 2.0, OpenID Connect, OAuth 2.0 and X.509v3). Additionally, it supports different endpoint services that the service providers are free to choose from, including B2ACCESS, EGI Check-in, eduTEAMS and INDIGO-IAM, which are described more in detail in Section 7.2.1. Furthermore, the EOSC-hub AAI includes a component for managing the access rights of the service's users, as well as a set of Token Translation Services (TTS) to translate between different protocols or technologies while passing identities and user roles to services.

Full documentation of the EOSC-Hub AAI is available at
https://confluence.egi.eu/display/EOSC/Authentication+and+Authorization+Infrastructure+-+AAI

5.3.2 *Availability monitoring services*

Availability monitoring allows service providers to get insights into the infrastructure, keeping an eye on the performance of their services to quickly detect issues and to help resolving them. The monitoring allows users and service providers to check the status of a EOSC service and to get information about its availability and reliability within a particular time period. The services also provide dashboard interfaces and allow sending real-time alerts.

The EOSC-hub service monitoring is based on the ARGO system[32]. The ARGO Service collects status results from one or more monitoring engines and delivers status results and/or monthly availability and reliability results of distributed services through a Web UI, with the ability for a user to drill-down from the site to the service to the individual result. The ARGO monitoring and alerting service provides the following features (more information in Section 7.2.2):

- Open source modular architecture
- Horizontal scalability
- Scalable elaboration of metrics

The integration with the EOSC monitoring services consists of requesting the monitoring of one/more services, opening a ticket via the Helpdesk (see Section 5.3.4), assigning it to 'EOSC Availability / Reliability Monitoring', and provide a short description of the integration use case, the name of the service and its endpoint(s), and the probes that should be used from the available ones[33] or to create new one(s).

5.3.3 *Accounting services*

The EOSC Accounting system is based on the EGI Accounting[34] system, described in more detail in Section 7.2.3, which can store, collect and aggregate user accounting records from various services, such as Cloud, HPC and storage usage. The usage data is collected from the respective Resource Centres that connect their service endpoints to the centrally managed Accounting Service. Data is then securely forwarded from the sensors into the central Accounting Repository where those data are

[32] https://eosc-hub.eu/support-services/Argo%20Service%20monitoring
[33] https://poem.egi.eu/ui/public_probes
[34] https://eosc-hub.eu/support-services/Accounting

processed to generate various summaries and views that can be consulted online through the EGI Accounting Portal[35], providing among others:

- More control over resource consumption
- Less requirement of defining data models, architecture, and setup of an accounting system
- Reduced cost of maintaining an accounting infrastructure
- Access to a reliable, high available, high performance service
- User friendly web interface

A service can be integrated with the Accounting Services either by i) registering the service in a 'topology', associating it with geographical or community entity (e.g., a country, a community); ii) installing the appropriate parsers at the service provider to produce accounting data in the format expected by the Accounting Repository[36]; iii) and sending directly or via some intermediate repositories the accounting records to the Accounting Repository.

5.3.4 *Helpdesk*

The EOSC-hub Helpdesk[37] is the central entry/contact point and ticketing system/request tracker for issues concerning EOSC services. New service providers can integrate into the Helpdesk leading to

- the creation of a corresponding support topic that is listed on the Helpdesk user interface (for users to ask questions or raise issues directly to the provider)
- the provider support team receiving notifications about tickets that are assigned to this topic by the users, or by the ticket handler team of EOSC-hub

The Helpdesk can provide the information and support needed to troubleshoot user's product and service-related problems and also, via the ticketing system, the users can report incidents, bugs or change requests. The helpdesk thus serves the following two groups:

- ***Users*** have the following features:
    - Creation of tickets for any EOSC services, display of all available tickets, notifications on the response of the tickets via the integrated ticketing system.
    - Login with the EOSC AAI system
- ***Helpdesk Team*** has the following facilities:
    - Notifications about newly created tickets, classification of tickets according to the priority level
    - First level of service for EOSC services
    - Interface with a Known Errors Database and with a Change Management Database
    - Management of incident or disruption of Hub services
    - Creation of a new support unit with assignation of an administrator role to specific users on request.

The Helpdesk can be integrated into EOSC services by the following ways:

- Using directly the EOSC helpdesk as the ticketing system for the service.
- Using the EOSC helpdesk only as a contact point to redirect the entry request for the specific service to a mailing list.
- Integrating an external ticketing system with the EOSC helpdesk infrastructure to enable transfer of tickets between the two.

In order to start the integration process, it is required to create a ticket in the Helpdesk with a subject including the integration way and the service name, and issue details including: web page(s) for the 0 Level (FAQ, How To, etc.) that users can use for solving issues without creating a ticket, the service's

[35] https://accounting.egi.eu/
[36] https://wiki.egi.eu/wiki/APEL/MessageFormat#Job_Records
[37] https://helpdesk.eosc-hub.eu/

responsible contact details, the SDT, and other information depending on the selected integration way.

5.4 Managing Research data

EOSC-hub offers a number of services for service providers to help them more easily manage research data and implement scenarios where research data need to be stored, transferred, analysed etc. A service can be benefited from such a support when the user's data needs to be stored and analysed using external computation or storage resources (e.g., cloud resources), or when the service needs an external repository to deposit research data or scientific applications for broader sharing and reuse. Figure 4 depicts relevant ESOC-hub services along the virtuous cycle of research.

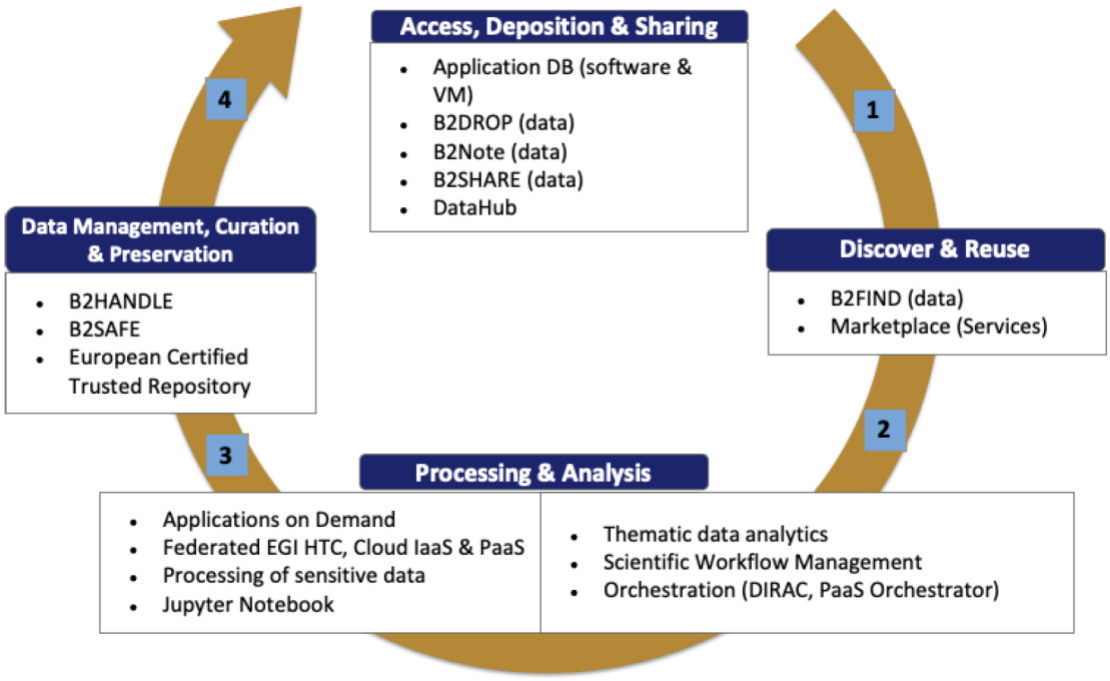

Figure 4: EOSC-hub services to support the management of research data

The EOSC services, available via the EOSC marketplace[38], can support different research data management tasks including:

- Helping in discovering and reusing existing research data and services or applications
- Processing and analysing the users' research data
- Archiving and curating research data
- Storing and sharing research data

Moreover EOSC-hub also provides consultancy service through the EOSC-hub Helpdesk that the scientific communities may choose in order to further uptake the data management services.

5.4.1 *Cloud Compute services*

Some of the research data management services supporting the data processing and analysis phase are those related to the computing and infrastructure resources that EOSC offers.

The EGI Cloud service [7], in particular, is implemented in the form of a Federated Cloud based on open cloud system federating institutional clouds in order to offer a scalable computing platform for

[38] https://marketplace.eosc-portal.eu/

data and/or compute driven applications and services. EGI Cloud Compute allows users to host data and compute intensive applications in a distributed infrastructure with complete control on the software and resources used. The main features provided by the EGI compute services are:

- Elastic computing infrastructure for the users, allowing the execution of compute and data intensive workloads, hosting of long-running services, or creation of disposable testing and development environments in VMs and containers
- VM image sharing and distribution allowing customized VM images to be easily shared to multiple clouds via the open 'Applications Database' library of Virtual Appliances
- Unified view of federation allowing single sign-on (SSO)[39] across resource providers, federated accounting over resource and service usage, federated monitoring to compute metrics for availability and reliability reporting, etc.
- More than VMs EGI, including support for Docker applications on EGI resources; allowing the use of the integrated PaaS and SaaS solutions, deployment of Hadoop, Docker Swarm etc. to access Object Storage and other IaaS capabilities

The EGI Federated Cloud operates as a federation of heterogeneous IaaS cloud services and enforces cloud technology agnosticism and service portability and service portability in a hybrid environment via the adoption of open standards.

The IaaS Cloud capabilities are integrated with the Image Management subsystem, provided as part of the Federated Cloud infrastructure. The EGI Federated Cloud currently integrates resources from OpenNebula[40], OpenStack[41] and Synnefo[42] providers. Figure 5 shows the schematic diagram of the EGI Federated Cloud Architecture. The cloud specific capabilities are: (i) Virtual Machine (VM) Management and Block storage management; (ii) Data Management, and (iii) Image Management. The cloud federation relies on services like:

- Federated AAI
- Federated Accounting
- Information System
- Federated Monitoring
- Federated Service Registry.

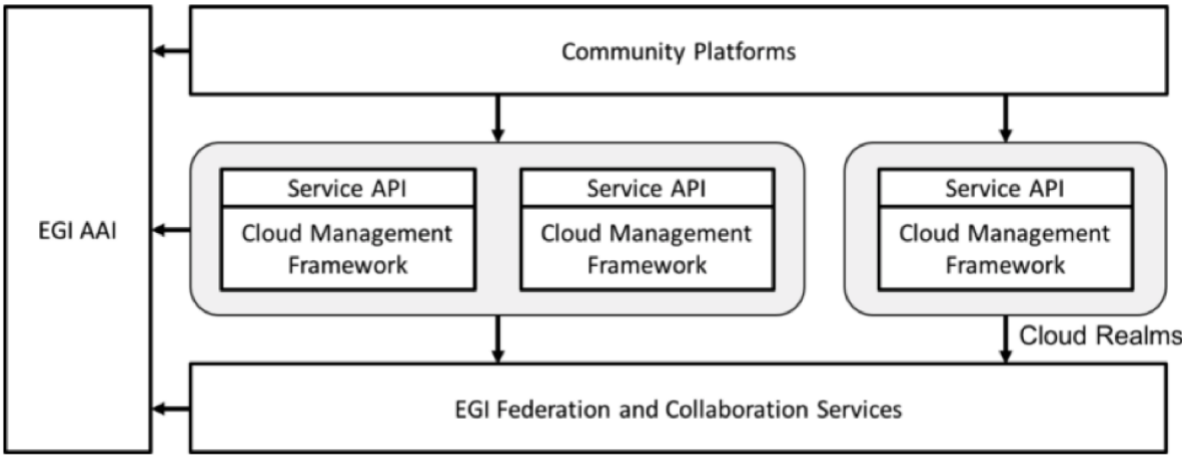

Figure 5: EGI Federated Cloud Architecture

[39] https://auth0.com/docs/sso
[40] https://opennebula.io/
[41] https://www.openstack.org/
[42] https://www.synnefo.org/

Cloud capabilities are provided by the individual cloud providers, while the AAI and the Federation and Collaboration Services are deployed to enable the federation. Users can choose to use a single cloud provider or multiple providers, depending on the locality of the data they need to process and the type of workflow to be executed. The EGI Federated Cloud Infrastructure as a Service (IaaS) resources centres deploy a Cloud Management Framework (CMF) enabling VM management, block and object storage.

The end-user capabilities are implemented via community agreed APIs that can be integrated with the following EGI services:

- AAI enabling SSO for authentication and authorization across the whole cloud federation
- Configuration Database to record information about the topology of the e-infrastructure.
- Federated Accounting to collect, aggregate and display usage information.
- Monitoring to perform federated service availability monitoring and reporting of the distributed cloud service endpoints, and to retrieve this information programmatically.

The EGI cloud federation currently deploys cloud sites from all across Europe. These clouds are available for users through community allocations, so called Virtual Organizations. Each Virtual Organization (VO) can get access to a subset of the federated cloud sites according to their local policies, and makes those available for the community members. VO members can deploy new VMs on the cloud sites through the EGI AppDB VM marketplace[43], and can instantiate VMs and block storages via the graphical AppDB VMOps Dashboard or using the API and command line interfaces offered by the cloud sites.

Few of the use cases since the deployment of the EGI Federated Cloud are:

- Service hosting to host any IT service as web servers, databases, etc.
- Compute and data intensive applications needing considerable amount of resources in term of computation and/or memory and/or I/O
- Dataset repository to store and manage large datasets
- Disposable and testing environments for training or testing new developments

5.5 Alignment with the service management system

EOSC-hub defines and implements the EOSC IT service management system[44] (ITSM), i.e., the activities performed by service providers to plan, deliver, operate and control services offered to customers. These activities are directed by policies and are structured and organized by processes and procedures.

The scope of the SMS is primarily all services contributing to creation and delivery of the hub. The hub is a set of services essential to provide the core functionality for EOSC like: helpdesk, monitoring, accounting, order management etc. The key benefits of the SMS are:

- To ensure robust and resilient service delivery of services within the hub to the EOSC federated infrastructure;
- To facilitate communication between customer and providers by introducing single point of contact (helpdesk, marketplace etc.);
- To disseminate and share service delivery best practices among providers;
- To facilitate alignment of service management activities of all of the service providers, supporting different levels of integration with the centralized services;

[43] https://appdb.egi.eu/
[44] https://www.eosc-hub.eu/key-exploitable-results/eosc-service-management-system-sms

- To integrate the services provided by the different providers into the common marketplace and monitoring frameworks in a way that provides value for EOSC.

The activities carried out in context of the SMS are structured and organized into processes and procedures according to the lightweight FitSM IT Management standard[45] that is aimed at facilitating service management in IT service provision.

There are 14 processes of FitSM that help service providers as follows:

- *Service portfolio management (SPM)* for defining and maintaining a service portfolio
- *Service level management (SLM)* to maintain a service catalogue, and to define, agree and monitor service levels with the customers
- *Service reporting management (SRM) to* specify and ensure that all service reports are produced according to specifications in a timely manner to support decision-making
- *Service availability and continuity management (SACM)* to ensure sufficient service availability to meet the requirements and adequate service continuity.
- *Capacity management (CAPM) to* ensure that sufficient capacities are provided to meet the agreed service capacity and performance requirements.
- *Information security management (ISM) to* manage information security effectively through all activities
- *Customer relationship management (CRM) to* establish and maintain a good relationship with customers receiving services
- *Supplier relationship management (SUPPM) to* maintain a healthy relationship with suppliers supporting the service provider in delivering services to customers, and to monitor their performance
- *Incident and service request management (ISRM) to* restore normal service operation within the agreed time after the occurrence of an incident and to respond to user service requests
- *Problem management (PM) to* investigate the root causes of (recurring) incidents in order to avoid future recurrence of incidents by resolving the underlying cause
- *Configuration management (CONFM) to* provide a logical model of all configuration items (CIs) and their relationships and dependencies
- *Change management (CHM) to* ensure changes to CIs are planned, approved, implemented and reviewed in a controlled manner
- *Release and deployment management (RDM) to* aggregate changes of one or more CIs to releases, so that these changes can be tested and deployed to the live environment together
- *Continual service improvement management (CSI)* to identify, prioritize, plan, implement and review improvements to services and service management

The impact of the EOSC-hub SMS process on the onboarded services depend on the choices the service providers make for integrating with other Hub Portfolio components (those described in Section 5.3). For example, onboarded services become in the scope of SPM when they are included into the EOSC Service Portfolio; enabling ordering to allow users to request access to a service via EOSC marketplace brings the scope of CRM into action; using the Helpdesk involves the ISRM process defining the timeline to reply to incidents; using AAI requires the provider to meet minimum security requirements and to accept EOSC security policy as defined in the ISM, and so on.

[45] https://www.fitsm.eu/

6 RELIANCE services to be integrated into EOSC

6.1 ROHub platform

ROHub[46] [8] is a holistic solution for the storage, lifecycle management and preservation of scientific investigations, campaigns and operational processes via research objects. It makes these resources available to others, allows to publish and release them through a DOI, and allows to discover and reuse pre-existing scientific knowledge. Built entirely around the research object concept and inspired by sustainable software management principles, ROHub is the reference platform implementing natively the full research object model and paradigm, which provides the backbone to a wealth of RO-centric applications and interfaces across different scientific communities.

ROHub can support different stakeholders, with the primary focus on scientists, researchers, students and enthusiasts, enabling them to manage and preserve their research work, to share it and make it available for publishing, to collaborate and to discover new knowledge. However, other user groups can be benefited by ROHub like the Industry that can leverage the platform to externalize their research to a community of researchers worldwide in multiple scientific domains, e.g., launching campaigns for research on specific topics, and then to follow and to monitor the progress. Similarly, investors can keep up to date and track scientific advances to fund and get involved in future breakthroughs. As another example, publishers can also leverage ROHub to advertise their journals with researchers, have access to a pool of potential reviewers, and implement more interactive, review processes.

ROHub is the result of continuous work on research objects since 2010 as part of Wf4Ever and then by EVER-EST projects and it is at the core of the RELIANCE portfolio. ROHub comprises a backend service (RODL), exposing a set of Restful APIs, a reference web client application (ROHub portal), and integrates multiple added-value research object services, as described below.

RELIANCE will leverage the extensive results produced during the journey of the research object and its supporting technology ROHub, since their conception, implementation and validation in experimental science in Wf4Ever project to their adoption and adaptation to the Earth Science domain in EVER-EST project. Some of the latest advances in this journey, which are available via ROHub, include: a new module extending the research object model to support the needs of scientists in Earth Science (e.g., geospatial metadata) and other disciplines (e.g., data access policies); a standard OpenSearch interface to facility integration with other repositories; the provisioning of digital object identifiers (DOI) to ROs to enable persistent identification and to give due credit to authors; the extension of the RO lifecycle to support forking mechanisms (e.g., to start new lines of research from an existing RO) that facilitates the reuse and attribution of research results; the generation of content-based, semantically rich, RO metadata extracted with natural language processing techniques, complemented with metadata aligned with standards used for datasets discovery, enhancing RO visibility, discoverability and reuse via search and recommendation systems and third-party search engines; various types of checklists that provide a compact representation of research object quality, along with an improved quality monitoring to prevent RO decay, as key enablers of scientific reuse; and complemented with mechanisms to associate subjective notion of quality about ROs and to keep account of its social impact (e.g., ratings, likes).

6.1.1 *ROHub APIs and added-value services*

The two primary APIs exposed by ROHub are the RO API and the Evolution API, which respectively define the formats and links used to: i) create and maintain research objects, the resources aggregated and the associated annotations (metadata); ii) change the lifecycle stage of a research object, create

[46] http://www.rohub.org/

an immutable copy (snapshot or archive) from a working (live) research object or fork it to start new line of work, and fetch their evolution provenance. ROHub exposes an OpenSearch compliant API that includes the Geo and Time extensions, facilitating the integration with other OpenSearch compliant catalogues. Other APIs include user management and access control.

Additionally, ROHub integrates the following key added-value services, each exposing a Restful API:

- RO enrichment service generates semantic machine-readable annotations (metadata) in the research object, based on its content (textual resources) through Expert System's Cogito system. This service is part of the Text Mining services stack, which will be explained in more detail in Section 6.3.
- RO search and recommendation service suggests research objects that might be of interest according to the user's research interests following a content-based approach, i.e., comparing the research object content with the user interest.
- RO notification service enables the subscription to events related to a particular research object (e.g., changes in content or quality), or to the repository itself (e.g., when new ROs are created).
- RO checklists service provides access to the minim-based checklist evaluation of research objects, used to assess their quality according to different purposes, e.g., completeness, runnability or repeatability. Each checklist captures the needs and expectations of a FAIR RO in a domain, or for a particular application.
- RO stability service enables the evaluation of the RO through time by capturing discrete values provided by the checklist service in different moments of its evolution. It allows testing the ability of a research object to achieve its original purpose through time, providing an indication of its stability and reliability.
- RO DOI service allows the provisioning of digital object identifiers to research objects to enable persistent identification and to give due credit to authors
- RO impact analytics service provides measures of the impact of a research object in a community, its citations and applications.
- RO social services enable the rating of research objects, marking a research object as favourite (like), and to comment it or its aggregated resources, providing a more subjective notion of quality about ROs (e.g., how it is perceived by other colleagues), and in turn of its impact in the community.
- RO paper generation service generates a paper view of the research object. The service creates a PDF file, using a conference paper template, automatically from the aggregated content and metadata.

Regarding the user authentication and authorization, the latest implementation of ROHub relies on a Keycloak solution for identity and access management. Keycloak provides a single sign-on solution, where users authenticate with Keycloak rather than individual applications. Moreover, Keycloak is based on standard protocols, providing support for OpenID Connect, OAuth 2.0, and SAML, and enables identity brokering and social login, thus managing the connection to external identity providers like Google, ORCID and others. This Keycloak solution will be the reference RELIANCE AAI, and it will be the service that will be connected to EOSC AAI.

6.1.2 *Aggregated resources and Data Cube-centric Research Objects*

Resources aggregated in a research object can be internal and/or external to the ROHub, provided that they are accessible via some URL. Thus, ROHub typically operates alongside personal storage services where scientists can easily save the auxiliary resources produced/consumed during the research lifecycle. RELIANCE will leverage this kind of facilities from EOSC offering, allowing it to support a larger number of scientists.

Resources aggregated in a research object can be of any type; however, depending on the main resources they aggregate and/or their purpose, research objects are classified into different kinds, e.g., workflow-centric, data-centric, bibliographic, etc. In RELIANCE, research objects will be extended to support Data Cubes as the mechanism enabling efficient access to large structured datasets. Data Cubes will be treated as first-class resources, aggregated and described in detail, e.g., how it was generated (specification) or used, in a Data Cube-centric research object. For instance, a research object might aggregate a Data Cube including its metadata like identification, description, resolution as well as the particular parameters used to generate the dataset or subset and the link for access. The new kind of research objects will support the realization of a FAIR research environment in EOSC, where not only the data resources are aggregated (among others), but also the mechanism to recreate, access and use efficiently large datasets.

Similarly, research objects can encapsulate any kind of methods or scientific workflows. In RELIANCE the focus will be on the exploitation of Jupyter notebooks, containing code, equations and other interactive content, as the processing and computing environment for research objects, leveraging EGI services.

6.1.3 *ROHub user interfaces*

The reference Web application for ROHub, called RO Portal, provides a comprehensive user interface exposing all the research object functionalities to the end-users (see Figure 6). Nevertheless, the APIs exposed by ROHub enable the implementation or extension of client applications that can leverage research objects as the mechanisms for research lifecycle management, thus enabling researchers to continue using their preferred environments and tools. ROHub services are already being used by applications like the EVER-EST VRE portals, an interactive web-based prototype application that integrates time series from UNAVCO and National Ecological Observatory Network (NEON) sensors, the collaboration spheres, the research object monitoring tool to visualize the RO quality and stability over time and to identify decay.

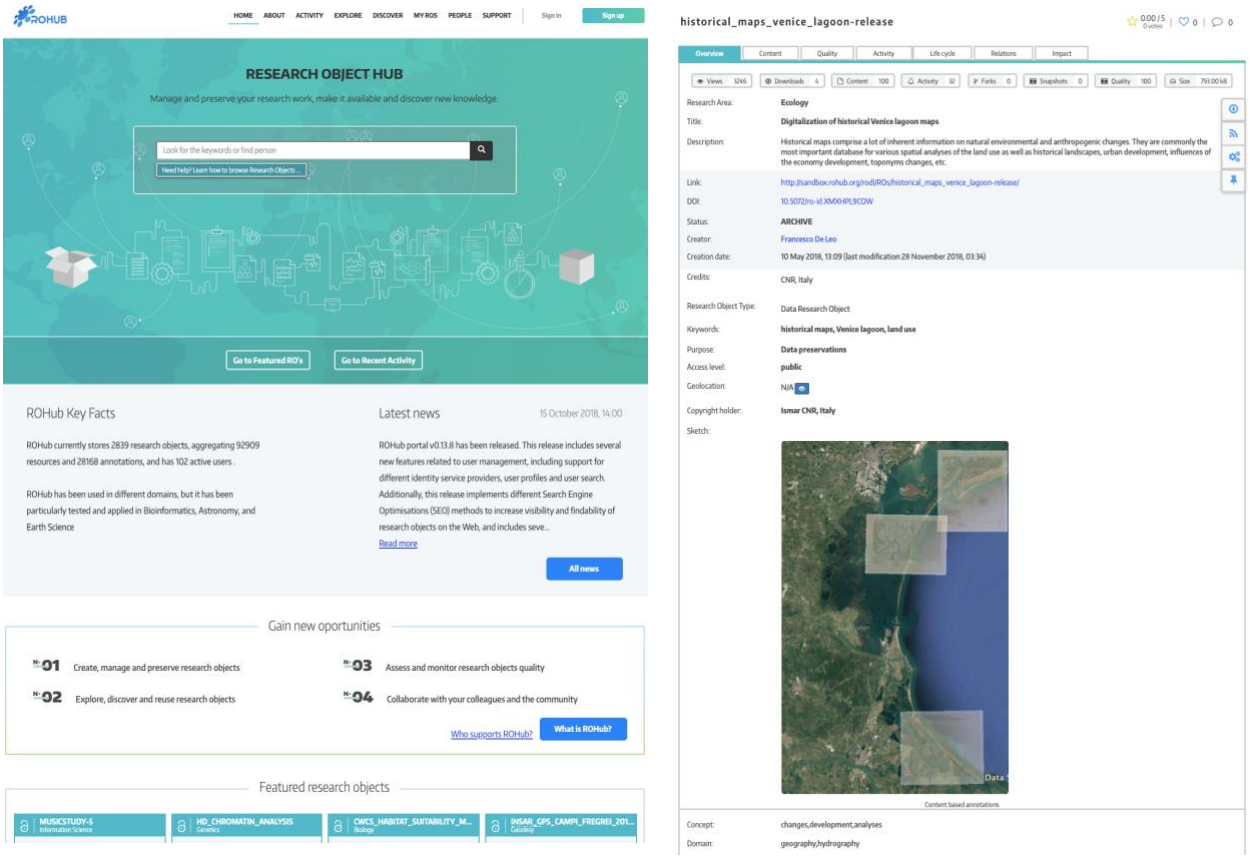

Figure 6: The RO-HUB Portal and a CNR Research Object example

6.1.4 *ROHub evolutions*

As part of the RELIANCE project ROHub will be extended with two added-value services, namely Data Cubes and text mining, which will be also part of the set of research enabling services of RELIANCE offering. These two services will enhance the capabilities of research objects to realize a FAIR research environment in EOSC, where not only the data resources are FAIRly managed via research objects (among others), but also the mechanism to recreate, access and use efficiently large structured datasets by treating Data Cubes as first-class entities. At the same time, textual (non-structured) resources (e.g., scientific papers) will be processed by text mining services in order to enrich the research objects with the metadata extracted automatically, thus increasing the FAIR-ness of these resources as well. These services, dealing with complementary resources (structured, non-structured), will enable research objects to bridge the gap between data services used in a research and the scholarly communication services used to publish the research results, connecting also all other resources used/produced in the middle with rich metadata.

6.2 ADAM platform

The Advanced geospatial Data Management platform (ADAM)[47] is a tool to access a large variety and volume of global environmental data. ADAM allows the extraction of global as well as local data, from the past, current time, as well as short term forecast and long-term projections. Most of the data is updated daily to allow users to always have the most recent data to play with.

[47] https://adamplatform.eu/

6.2.1 *Principles and architecture*

The core of ADAM is a Data Access System (DAS), a software module that manages a large variety of geospatial information that features different data formats, geographic / geometric and time resolution. It allows discovering, accessing, visualizing, sub-setting, combining, processing, downloading all data sources simultaneously. The only requirement is that each dataset shall feature position and time tags. The DAS exposes OGC Open Search, Web Map Service and Web Coverage Service (WCS 2.x) interfaces that allow discovering available datasets and subset them in any dimension with a single query.

ADAM is a modular platform: various DAS are deployed on different data sources (DIAS Mundi, DIAS creodias, Amazon Web Services - AWS, MEEO Data Facility, SISTEMA Data facility), allowing accessing and sub-setting the available datasets without downloading / duplicating the data. Distributed data sources are made accessible through the data cube layer that exposes OGC-standardized interfaces. On top of the data cube layer, platform-based interfaces (web application, mobile application, Jupyter Notebook and APIs) as well as third party user interfaces can be deployed (see Figure 7 below).

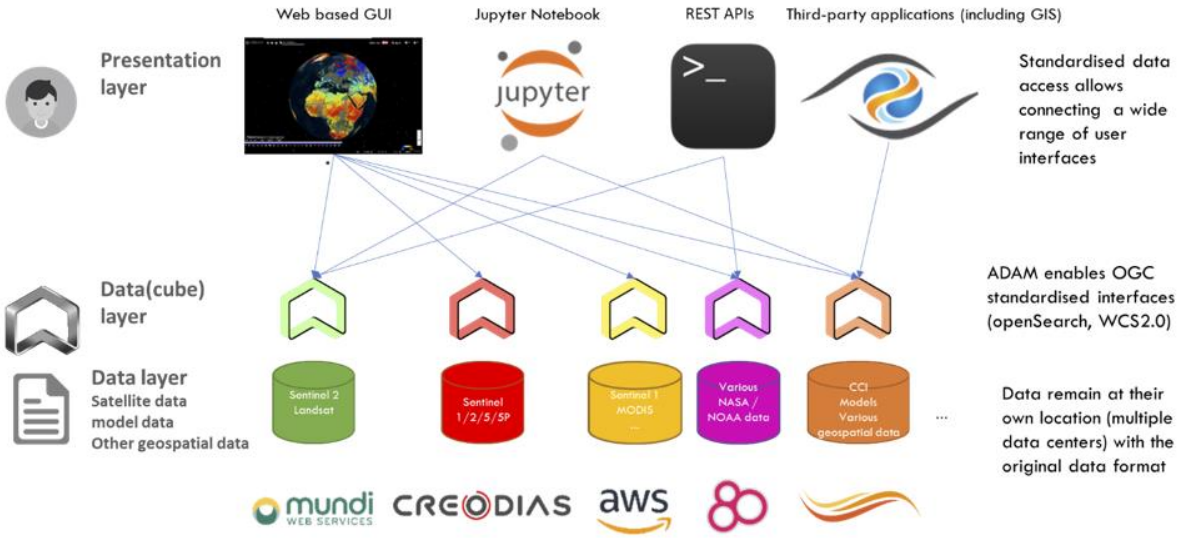

Figure 7: ADAM high-level architecture

ADAM, referred also as a virtual data-cube technology, enables seamless access to all available datasets through a single interface to allow developing user-centred solutions on top of it; cutting the access barriers to a large variety of data (no need to know the exact format of any of the available datasets) it saves data preparation time and facilitates the simultaneous use of data coming from different sources

ADAM offers the following interfaces for data discovery and access:

- the Explorer, a web-based graphic user interface to allow users to explore, access, process and download data. Explorer includes also and Operator Interface, data processing interface and a mobile application (ADAM Mobile)
- The Application Processing Interfaces (APIs), that provide a python-library to directly access the ADAM data access and processing capabilities directly integrated in the user's code and applications.
- the Jupyter Notebook, a web-based processing environment to allow users to import, write and execute code that runs close to the data, exploiting the power and the APIs on a remote computation environment (no user resources are used)

6.2.2 *Authentication and Users Management*

The Authentication, Authorization and Accounting (AAA) service is a generic module based on OAuth and OpenID Connect (OIDC) technologies, that allows the integration of customized as well as third-party authentication systems, such as NextGEOS SSO, Google, LinkedIn, and any other OAuth/OIDC provider. It can also be adapted to manage Microsoft Active Directory login.

The module supports:

- Authentication, the process of identifying an individual;
- Authorization, the process of granting or denying access to various resources based on the user's identity;
- Account provisioning, the process of providing users with access to data and processing resources, like OpenSearch and CSW catalogues, WCS host and products, WMS endpoints, and Pipelines;
- Users and Groups administration, the process of creating new and modifying or deleting existing identities as well as managing the security entitlements associated with those identities.

6.2.3 *ADAM explorer*

The Explorer[48] allows, after authentication, the user to perform the following main operations:

- Dataset discovery, from the “Datasets' catalogue” the user can discover the available datasets and add to his/her own “Dataset basket”
- Dataset exploration, from the “Dataset basket” the user can load in to the system and explore spatial and temporal coverage of his/her dataset of interest
- Dataset access, a series of options are available to retrieve data over area and point of interest

The Figure 8 shows snapshots of the ADAM Explorer instance for RELIANCE: Global Land Surface Temperature from MODIS on March 23rd 2021 and time series over Ferrara, Madrid and Poznan cities in a 30 days period are presented on the desktop view.

Figure 8: ADAM explorer showing Global Land Surface Temperature from MODIS on March 23rd 2021 and timeseries over Ferrara, Madrid and Poznan cities in a 30 days period

[48] https://explorer.adamplatform.eu/

6.2.4 *ADAM APIs and OGC services*

The ADAM API exposes the data discovery and access service interfaces in a standard python package called adamapi. To use the ADAM API, the user must have an "ADAM API KEY", the authentication key linked to the ADAM instance. Currently the adamapi package provides following functions:

- Auth, the module to configure the environmental variables and a method to manage the authorization
- Datasets, the module to discover the datasets available in the ADAM instance, including all properties (e.g., description, start/end date, spatial coverage)
- Search, the module to discover the products available for a specific dataset, including filter options by supported parameters (e.g., geometry, attributes, tile, …).
- GetData, the module to access a product or a timeseries of products, up to pixel granularity, including the support of different encodings (json, tiff, png, gif).

Additionally, ADAM exposes standardized Open Geospatial Consortium (OGC) services for data discovery, data visualization, access and data processing, to facilitate the integration of ADAM resources within user's stacks.

- Web Map Services (WMS) is available to users who want to quickly visualize datasets within their clients.
- An OpenSearch catalogue service (CSW) is exposed to discover all data within the platform through the various data facilities. The connection to INSPIRE compliant catalogues is also supported.
- The Web Coverage Service (WCS) is the core part of the DAS module. WCS can be queried directly via REST calls. Authentication as well as quota management is supported via token released by the AAA module

6.2.5 *Jupyter notebooks*

The Jupyter notebook[49] is an interface offered by ADAM to users willing to exploit their own algorithms or to develop new code directly on the platform. The Jupyter Notebook is an open-source web application that allows users to create and share documents that contain live code, equations, visualizations and explanatory text. ADAM offers a python console and platform libraries (data access, visualization, processing functions) to give the users the maximum flexibility for data exploitation. In order to facilitate the usage of the Jupyter notebook, examples are provided for each functionality pre-loaded for each user. In case users will need further libraries and tools (e.g., rsgis lib) they can be deployed in all notebooks taking advantage of the data offer and functionalities of the user research environment.

6.2.6 *ADAM evolutions*

As part of RELIANCE, ADAM will be interconnected with ROHub, e.g., enabling a single sign-on between the services, as well as enabling researchers to open a data-cube centric research object in the ADAM interface, or to save back data cubes in a research object.

With respect to the existing data cubes offer, new data cubes need to be generated to accommodate RELIANCE communities' requirements: data processing pipelines to generate analysis-ready data and enable pixel-based access services will be implemented as part of the ADAM evolutions.

Additionally, ADAM plan to leverage and reuse existing EOSC services, including EOSC AAI (via RELIANCE AAI), but also potentially storage space for researchers to store their data cubes (B2DROP),

[49] https://jupyter.adamplatform.eu/

or computing power to manipulate heavy data cubes, and especially the Notebooks services that will allow supporting a larger number of users and research communities.

6.3 Text Mining services

We plan to integrate in EOSC text mining and analytics services that helps user communities to tap into the information encoded in text that is produced across all the stages in the scientific endeavour. These services transform text descriptions coming from research resources into structured data. Examples of research resources are datasets, data cubes, research objects, code repositories, collaborators, news and scientific papers. The structured data extracted from research resources is semantic metadata that describe the content of the text beyond the traditional keywords and text snippets, including named entities, multiword expressions, and concepts. Semantic metadata is a key enabler of the FAIR principles since it helps to increase the findability of the research resource, thus fostering their sharing and reuse. In addition, semantic metadata can be used to deliver more accurate search results and content-based recommendations, and fuel analytic services that can support the scientific enterprise.

6.3.1 *Information extraction and semantic annotation service*

The goal of the information extraction and semantic annotation service is to generate structured metadata from the textual content of the research resources, that will be used to enrich the source of such text, both individual documents and aggregations of documents, with semantic metadata. This metadata will help to synthetize high level concepts from the research resources, that therefore will increase its chances to be found and reused by other researchers.

The methodology used to extract textual information depends on the type of research resource that is being analysed. In the case of PDF documents, the textual content from these files is extracted by open-source tools such as apache PDFBOX[50], while we will use apache POI[51] to process Word documents and PowerPoint presentations. Regarding datasets and data cubes, we will need to extract the text from README files and those textual pieces that describe the features of their content. Code repositories and Jupyter Notebooks must be parsed to select text information from comments and markdowns texts. Research objects, as heterogeneous information container units, will be processed to extract all the possible pieces of textual information from them. Furthermore, the potential metadata that scientific papers host, such as titles, abstracts, authors, tables, figures, and citations, which appear in the text of any scientific publication but are rarely part of the structured information hosted in scientific publication repositories, will be extracted using tools as GROBID[52], a machine learning library for extracting, parsing, and re-structuring technical and scientific PDF documents. We will also assess the feasibility of extracting data cubes references from this type of texts, enabling the access to datasets which may be relevant for the research discussed in a scientific publication or hosted by a research object.

Once the pieces of text are extracted, they are fed into expert.ai[53] Natural Language API (NL API herein) to generate representative metadata from the text content of the research object. The NL API has two main components: the sensigrafo (a semantic network where knowledge is represented as a graph of concepts and relationships between them), and the disambiguator (a multi-level linguistic engine able to disambiguate the meaning of a word by recognizing its context). The NL API uses both tools to perform a semantic analysis that disambiguates the meaning of the words from a document, associating tokens with concepts, and providing structure to the document. In addition, the NL API

[50] https://pdfbox.apache.org/
[51] https://poi.apache.org/
[52] https://grobid.readthedocs.io/en/latest/
[53] https://www.expert.ai

has a Named Entity Recognition module that aims at spotting names referenced by contiguous spans of tokens. These names are categorized in names of People, Organizations, and Locations. These entities, added to the concept structure provided by the NL API support formal text processing tasks such as indexing, classification, summarization, and translation. Therefore, the text extracted from the research object resources is processed by the NL API, generating the following metadata:

- Main Concepts: Most frequent sensigrafo concepts mentioned in the text.
- Main Domains: Fields of knowledge in which the main concepts are most used.
- Main Lemmas: Most frequent lemmas found in the text.
- Main Compound Terms: Most relevant phrases including multiword expressions found in the text.
- Named Entities: all the named entities found in the text classified into People, Organizations and Places.

In addition to the semantic metadata, we can include some text classification, based on the ANZSRC's Field of Research taxonomy. The Australian and New Zealand Standard Research Classification (ANZSRC), and more specifically, the Field of Research (FoR) classification is a taxonomy initially released in 2008, and updated in 2020, that allows technical and scientific documents to be categorized according to their field of research. It comprises 22 first level categories, which in turn have second and third level categories.

6.3.2 *Content-based retrieval and recommendation services*

Finding relevant research resources is a key step in the exploration stage of research activities. Usually, researchers rely on general-purpose search engines and specialized search engines, e.g., on scholarly communications or research objects, to find research resources. Nevertheless, general purpose search engines retrieve a wide set of results that must be inspected and filtered by users, and possibly requiring to refine the search by adding or removing keywords. Specialized search engines, on the other hand, focus on a type of research resources such as papers, or research objects. Therefore, often researchers need to use several search engines to find what they are looking for.

In RELIANCE we propose a search engine focused on retrieving only relevant resources for the scientific enterprise, thus avoiding the need to jump from one search engine to another. This search engine leverages the metadata extracted with the help of the information extraction and semantic annotation service to deliver more accurate results. In addition, in RELIANCE we complement the search engine with a recommendation service that suggests scientific resources related to one or more resources of interest for the researcher, leveraging also the metadata generated with the information extraction and semantic annotation service.

To implement the search engine and the recommendation system we are considering OpenAire services as a source of research resources. OpenAIRE is a technical infrastructure that harvest research outcomes from connected data providers. We plan to leverage the OpenAire Research Graph[54] that connects publications, datasets, software, and other research products.

6.3.2.1 Free text search service

This service retrieves research resources based on user queries. The search leverages the textual descriptions of the research resources and the semantic metadata added by the information extraction and semantic annotation service to these resources. The use of semantic metadata in the search engine allows searching for research resources sharing concepts and named entities. As opposed to keyword-based search engines where only documents containing a given keyword are retrieved, semantic search engines retrieve documents where a given concept was mentioned regardless of the keyword used to refer to the concept. In addition, named entities support more

[54] https://graph.openaire.eu/

accurate searches since users can specify the type of the entity they are looking for (e.g., person, location, organization) and the name of the entity (e.g., Paris). The search engine is built on top of a document index such as Solr[55] or ElasticSearch[56]. These platforms offer fast indexing of text documents and response to user queries at a scale. We index both the text from research resources and the structured metadata added by the information extraction service. This allows users to query for documents using keywords, as well as specific types of information such as concepts or named entities.

6.3.2.2 Content-based recommendation

Content and social-based recommendation systems could help to maximize the share and reuse of scientific publications and resources, encouraging connections between researchers and supporting community collaboration. Given the textual content from a research resource, or from the defined interest profile of a researcher, the recommender system suggests related resources that might be of interest. The semantic metadata generated from a research resource is leveraged when selecting the resources of the recommendation, helping with the comparison of two resources in terms of the concepts mentioned in them instead of only using plain words. The social dimension is exploited when defining the recommendation content and includes co-authors and researchers that share common interest. This defined profile of a researcher can be generated using the textual content of their own publications and the type of resources that is being used by them.

The Collaboration Spheres CS web application[57], is the user interface of the recommender system (see Figure 9). The user interface consists of a navigation panel on the left-hand side, the spheres on the right-hand side, the authentication box on the upper right corner and the help option just below this box. Broadly the navigation panel is where the user searches for the research resources or users to be included in the recommendation context. The spheres component serves as a container for the recommendation context and the recommender results. It consists of three concentric circles which show the level of relatedness of the recommendation. The central circle contains the recommendation target, which can be one or more research resources or researchers, while the recommended items are displayed around the outer circles. The closer they are to the central point, the strongest is the level of relatedness between the recommended resource and the target elements. Finally, the information of any recommendation will be displayed just by clicking on the item displayed on the spheres.

[55] https://solr.apache.org/
[56] https://www.elastic.co/elasticsearch/
[57] http://everest.expertsystemlab.com/spheres

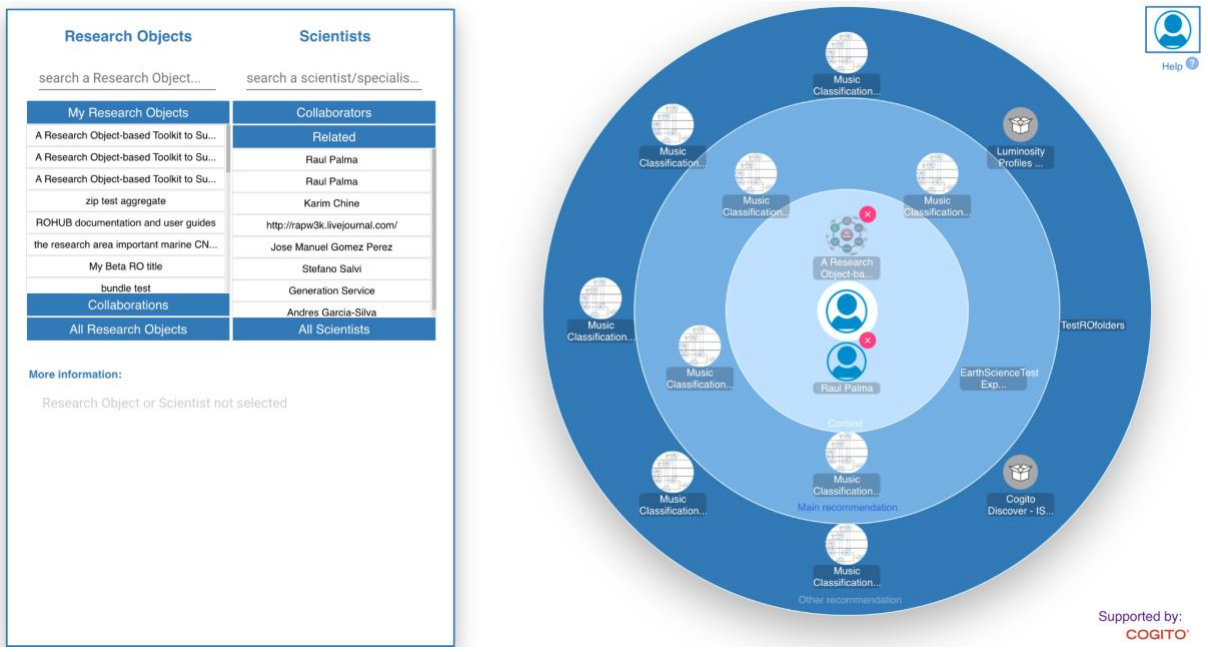

Figure 9: Collaboration Spheres Web application

6.3.3 *Extended analytics services in support of the scientific enterprises*

We plan to integrate in EOSC added value services for researchers and their communities leveraging the infrastructure on which the content-based retrieval services are built. Such services include accounting for the impact of a research work and how that influence propagates across researchers, and estimating the innovation potential of a research in terms of what has been already done in the current state of the art. These services are experimental, so we expect to deliver them as proof-of-concepts. Note that the list of services presented herein might be slightly modified according to the user communities needs to address the case studies in RELIANCE. In addition, we will analyse existing services in OpenAIRE catalogue to avoid duplicating services already available.

6.3.3.1 Influence Network

This service identifies the influence network of a research work showing its impact, highlighting the number of its citations and materializing the influence of the research in the form of an explicit graph of researchers and their work. The deployment of this service depends on the availability of citation data in OpenAIRE.

6.3.3.2 Novelty score

This service measures the innovation potential of a research work. The novelty score is based on the similarity of the analysed research paper with available papers in the literature. The intuition is that similar research to previous work with a certain impact in the community is unlikely to be innovative. The novelty score ranges between 0 for non-novel ideas to 1 for novel ideas. To implement this service, we fine-tune pre-trained language models, implemented using the transformer architecture, on a text similarity task. Transformer architectures are the current state of the art for many NLP tasks[58]. The most well-known transformer for NLP is BERT (Bidirectional Encoder Representations from Transformers) that was presented by Google research in 2018. Since then, many new variations of BERT such as DistilBERT and RoBERTa have been proposed pushing the state of the art even

[58] See GLUE benchmark leaderboard: https://gluebenchmark.com/

forward. The main benefit of these models is that first they are pre-trained on a large corpus, and then can be fine-tune for a particular task requiring few training data and computational resources

6.3.3.3 Support to reading comprehension

Understanding a scientific text implies capturing relevant information from it, processing concepts and ideas, and incorporating them into your knowledge base. Thus, it ultimately means to gain capabilities in answering questions related to the topics that are discussed within the text. Generating questions from a scientific publication could help to synthesize and support the knowledge that is derived from the text. Therefore, we will study the possibilities of generating a hierarchy of question-answer pairs related to the paragraphs from a scientific publication and producing a checklist of questions based on the publication that will assure its comprehension by a third-party researcher reading it.

6.3.3.4 Text mining and enrichment dashboard

We provide researchers with a dashboard that allows a simple and easy way to visualize in one glance the information extracted from scientific text by the mining and enrichment services. This dashboard allows researchers to search for research resources, visualize the results, and filter them using the semantic metadata. The dashboard has several interactive widgets to show aggregated data such as histograms, line and bar charts, pies, time series, and maps, among other visualizations. To implement the dashboard, we can use either Banana[59] if we work with a Solr index in the search service, or Kibana[60] in case we use ElasticSearch. These kinds of dashboards are referred to as descriptive analytics tools that are used to support understanding the data by visualizing it in an interactive way. The end users' involvement is required to explore the data, understand it, and make decisions based on the information

6.4 Copernicus Earth Observation Data Pipelines

Copernicus Earth Observation Data Pipelines is an integrated intermediate service layer for the systematic execution of Earth Observation Applications continuously delivering data and information to different research users. A data pipeline is the solution provided to a specific data challenge defined by a researcher as a tailored data processing workflow responsible for information extraction from a wide range of large volume data sources that are executed within defined spatial and temporal intervals.

6.4.1 *Background*

Earth Observation (EO) refers to the use of remote sensing technologies to monitor land, marine (seas, rivers, lakes) and atmosphere. Satellite-based EO gathers imaging data through satellite-mounted payloads that are then processed and analysed in order to extract different types of information. These payloads can contain optical, thermal and radar sensors.

Copernicus is the European Union's EO programme designed to meet the needs for space-derived, timely and accurate geospatial information. There are currently seven missions under the Copernicus Sentinel programme (Sentinel 1, 2, 3, 4, 5P, 5, 6) and also six thematic services (Land, Marine, Atmosphere, Climate, Emergency and Security) supporting the development of many applications. The Copernicus services process and analyse the data, integrate it with other sources, offering Geo-Information Systems (GIS) products to their users, and serving public authorities and commercial businesses.

[59] https://github.com/lucidworks/banana
[60] https://www.elastic.co/kibana

In 2017, the European Commission launched an initiative to develop the Copernicus Data and Information Access Services (DIAS) that facilitate access to the data of the Sentinel missions and information from the Copernicus services. Four consortia were chosen to set up DIAS computing environments under ESA management, and a fifth consortium is managed by EUMETSAT, ECMWF and Mercator Ocean International (MOI).

The DIAS represented an opportunity to federate the access to the Copernicus data and information close to processing facilities allowing further value extraction from the data, and respond through a dedicated service approach that is complementary to traditional data downloading. The DIAS targeted access to Copernicus data and information close to processing facilities and, through this, created the possibility to easily build applications and offer added-value services. In June 2018, the five DIAS entered their operational phases, acquiring customers and developing their business in a commercial environment. Nevertheless, after the end of the procurement contracts the operational continuity of the platforms and established services, customer relations and innovative solutions might be at risk.

Integrating Copernicus data and services into the EOSC service catalogue would facilitate access to the data of the Sentinel missions and information from the Copernicus services. By federating existing research infrastructures and scientific clouds the EOSC offers an open pan-European virtual environment to access, store, analyse and re-use data and support the development of cloud-based services for open science.

The EOSC can offer an integrated virtual environment to deposit, share and reuse EO data and could provide the computing resources necessary to manipulate and manage large amounts of Copernicus data. It will be easier to retrieve and process EO data, and this will stimulate demand for EO services and market opportunities for EO data services. The primary beneficiary of the EOSC would be Europe's research and scientific community, but the EOSC aims to widen its access and support public services, industry and ultimately society as a whole.

Even if users of EO data are present in many industries, each with very specific needs, the raw data coming from Copernicus missions cannot be exploited directly. Specific EO applications derive information contained in the images and act as the interface between the satellite technical features and the end users' specific needs.

The availability of the free and open Copernicus Sentinel products, together with the availability of EOSC computing resources, creates an opportunity for the wide adoption and use of EO data in a growing number of research fields. The latest advances in scientific data management provide easy and seamless access to the relevant data repositories as well as efficient operations (search, retrieval, processing/reprocessing, projection, visualization and analysis) to extract and distribute single parameters or combined products needed by researchers.

6.4.2 *Pipelines overview*

Copernicus EO data pipelines is an integrated EO intermediate service layer for the systematic execution of EO Applications to continuously deliver data and information to different users, complying with their specifications for information content and format.

In its essence, a data pipeline is the solution provided to a specific data challenge defined by a researcher. It is a tailored data processing workflow responsible for information extraction from a wide range of large volume data sources that are executed within defined spatial and temporal intervals and publishes the generated results. An example of a data pipeline is, for instance, the continuous production of orthorectified and atmospherically corrected Sentinel 2 vegetation indexes or providing Sentinel-1 based land deformation or change detection maps for a given area.

The EO Applications executed in the Data Pipeline are specific data processing functions defined by the researcher that perform data operations like processing / reprocessing, projection, visualization or analysis. The applications can be written in a variety of coding languages (e.g., Python, R, Java, C++, C#) and make use of specific software libraries (e.g., SNAP, GDAL, Orfeo Toolbox). Each data processing function is an application (e.g., a command line tool) that is a non-interactive executable program that reads some input (or a set of inputs organized in an atomic unit), performs a computation, and terminates after producing some output. Each application is released, packaged and built to be deployed in an IaaS provider ready to process inputs of the data pipeline.

The Data Pipeline uses shared data processing infrastructures where the applications are deployed, providing a hybrid and reliable cloud infrastructure for the applications needs. The service manages the IaaS and Cloud resources, orchestrates the execution of the triggered Data Pipelines and publishes the results. Working with Cloud-hosted data, the data pipeline provides fast access to the data sources and processing resources. Together with the availability of interfaces, APIs and tools for accessing and storing the data, Cloud computing provides the on-demand delivery of computing power, servers, databases, networking, software, analytics and other resources that support the data pipeline operations.

The Data Pipelines are triggered in two modes:

- The data driven approach queries an associated EO catalogue and retrieves the list of data products to be processed.
- In the Event Driven approach, there is an external service that announces an event (e.g., earthquakes, volcanos) and generates a spatial & temporal domain of execution. For the data stage-in the input products are retrieved and made available as inputs for the local processing of the EO Application. For the data stage-out, the outputs generated by the processing are retrieved and automatically published onto an external persistent storage together with the generated metadata.

According to the need identified by the researchers, the stream of outputs of a data pipeline are delivered in an agreed format (e.g., CoG, zarray) and ready to be consumed through a Jupyter Notebook or to be ingested in a dedicated Data Cube

In RELIANCE, we plan to leverage EOSC resources for the EO data pipelines, particularly those services related to cloud compute delivering federated IaaS and PaaS, as well as access to federated storage. These would enable the creation and execution of Copernicus EO data pipelines at scale, supporting a large number of users and research communities.

7 RELIANCE and other EOSC services

This section discusses existing EOSC services that could be used and integrated with the RELIANCE services. These services have been selected based on integration requirements of EOSC, and based on the specific needs of the project services.

In addition to onboard the RELIANCE services (described in Section 6) into the EOSC (i.e., publishing them in the EOSC Portal), we plan to expand their capabilities or scale them by reusing some of the existing EOSC federation services, research data management services and/or compute services. The next subsections provide an overview of these specific services.

7.1 EOSC portal

RELIANCE is one of the projects funded under the INFRAEOSC-07-2020 programme focused on increasing the service offer of the EOSC Portal. In this line, the onboarding and registration of RELIANCE services in the EOSC portal is one its core tasks.

As described in Section 5.2, the EOSC Portal is the universal access channel to EOSC services and resources. Through the portal, researchers and professionals can discover and access open and seamless services, data, and other resources from a wide range of national, regional and institutional public research infrastructures across Europe, including computing, storage, data management, networking, research publications and software.

By onboarding RELIANCE services into EOSC portal, we will be able to get providers benefits, such as:

- advertise them on the EOSC Portal and promote their adoption outside our own user communities, reaching a wider user base.
- get statistics about access requests and customer feedback
- get a free online platform where it is possible to manage service requests, interact with users and provide them support, and agree the most suitable service levels.
- allow users to authenticate with their own credentials to access the services and resources and get support to enable this.
- contribute to the definition and maintenance of the EOSC service provisioning policies and the portfolio roadmap.
- join the group of providers that meet EOSC quality standards.

The universal entry point to the EOSC services is known as the EOSC Portal Catalogue and Marketplace, which is the result of collaboration and merge of eInfraCentral Service Catalogue and EOSC-Hub Marketplace. The resources in the catalogue (currently over 280), are grouped in the 8 categories including: networking, compute, storage, sharing & discovery, data management, processing & analysis, security & operations, training & support. RELIANCE will bring its research enabling services into this catalogue following the onboarding process described in Section 5.2.

7.2 Federation services

7.2.1 *AAI services*

RELIANCE research enabling services require user authentication and authorization. The services rely mainly on external identity providers (IdPs) for authenticating their users. ROHub, for example, supports social media identity providers like Google, academic identity providers like ORCID, community identity providers (e.g., EVER-EST project AAI) and a local identity provider. Similarly, ADAM platform supports google in addition to a local and a community identity provider. However, as part of RELIANCE both platforms will support single sign-on and they will support authentication via EOSC AAI services.

In particular, the latest implementation of ROHub relies on a Keycloak solution for identity and access management, which manages the connection to external identity providers. This Keycloak instance will, thus, act as the reference AAI in RELIANCE, and is the service that will act as a connection point between the RELIANCE services and the EOSC AAI. The EOSC AAI service itself will then act as a bridge between RELIANCE services and a broad set of identity providers, enabling users to login to the services from a portfolio of endpoints.

The EOSC-hub AAI comprises different AAI services, namely, EGI Check-in, eduTEAMS, B2ACCESS and INDIGO-IAM, described below. Resource providers can leverage these services for managing their users and their respective roles and other authorization-related information. Hence, in RELIANCE we had to choose which of these services to use:

- EGI Check-in[61] is a proxy service that operates as a central hub to connect federated IdPs with EGI Service Providers. Check-in allows users to select their preferred IdP so that they can access and use EGI services in a uniform and easy way. Main characteristics:
    - Enables multiple federated authentication sources using different technologies
    - Increased productivity and security
    - Federated in eduGAIN as a service provider, publishing REFEDS RnS and Sirtfi compliance
    - User registration portal to allow accounts-linking
    - Combines user attributes originating from various authoritative sources (IdPs and attribute provider services) and delivers them to the connected EGI Service Providers in a transparent way.
- eduTEAMS[62] enables members of the research and education community to create and manage virtual teams and securely access and share common resources and services using federated identities from eduGAIN[63] and trusted IdPs. For community managers, eduTEAMS provides a central point for the community to manage its user membership, to connect IdPs and Service Providers and to define and apply access and sharing policies, enable secure access and sharing of common resources and services, and manage groups and roles, while for users it provides:
    - Sign in to services with existing identities via eduTEAMS
    - First class support of eduGAIN Identity Providers
    - Support for the research and scholarship entity category, code of conduct and Sirtfi
    - Support for a wide range of external IdP, such as ORCID and Google
    - Support for Web and non-Web based services - access to HTTP APIs
    - Account linking
- B2ACCESS[64] is the EUDAT federated cross-infrastructure AAI framework for user identification and community-defined access control enforcement. It provides arbitrating access to registered Service Providers via different protocols. Integrated Service Providers consume Attribute assertions from the service when the End User accesses them. B2ACCESS allows users to authenticate themselves using a variety of other IdPs or credentials. Its main features include:
    - Support of several methods of authentication via the users' primary identity providers (OpenID, SAML, x.509)
    - Can be used as primary identity provider

[61] https://www.egi.eu/services/check-in/
[62] https://eduteams.org/
[63] https://edugain.org/
[64] https://eudat.eu/catalogue/B2ACCESS

    - Can be integrated with any service of the EDUAT CDI[65] service provider federation
    - Is integrated with eduGAIN and supports identities from theoretically hundreds of Universities and Research institutions worldwide.
    - Provides unique and persistent EUDAT-wide meaningful identifiers.
- INDIGO Identity and Access Management (IAM)[66] service provides a layer where identities, enrolment, group membership and other attributes and authorization policies on distributed resources can be managed in a homogeneous way, supporting identity federations and other authentication mechanisms (X.509 certificates and social logins). The service has been integrated with many components like Openstack, Kubernetes, JIRA and Confluence, Grafana and with key Grid computing middleware services (FTS, dCache, StoRM). Main features include:
    - Authentication: via SAML IdPs or identity federations, OpenID Connect providers and X.509 certificates.
    - Enrolment: enrolment and registration, so that users can join groups/collaborations according to well-defined flows.
    - Attribute and identity management: manage group membership, attributes assignment and account linking.
    - User provisioning: the IAM provides endpoints to provision information about users' identities to other services.

In RELIANCE, our plan is to leverage EGI check, which also federates eduTEAMS.

7.2.2 *Availability monitoring*

Availability monitoring, as introduced in Section 5.3.2, is a service where users and service providers can check the status of an EOSC service and get information about its availability and reliability in the selected time period. In the case of RELIANCE, ROHub for example, uses Sentry[67] for such purpose. Sentry is a cloud-based error monitoring that helps software teams discover, triage, and prioritize errors in real-time. Additionally, ROHub plans also to use Nagios[68], which offers monitoring and alerting services for servers, switches, applications and services.

EOSC offers different monitoring services as described below, and which can be potentially utilized by RELIANCE:

- EGI Service Monitoring[69] keeps an eye on the performance of IT services and detects and resolves any issues. The service monitors the infrastructure by collecting the monitoring data generated by functional probes. The raw data is merged into statistics and available through the user interface in a user-friendly way. The service is based on ARGO[70], a lightweight service for Service Level Monitoring designed for medium and large sized Research Infrastructures. The main features include:
    - Minimal development effort for setting up monitoring services
    - Ready-to-use user interface
    - Automated reporting tools
- PerfSONAR[71] is an open-source, modular and flexible architecture for active network performance monitoring that provides a view of network performance across multiple

[65] https://eudat.eu/eudat-cdi
[66] https://indigo-iam.github.io/docs/v/current/about.html
[67] https://sentry.io/
[68] https://www.nagios.org/
[69] https://www.egi.eu/internal-services/service-monitoring/
[70] http://argoeu.github.io/overview/
[71] https://marketplace.eosc-portal.eu/services/perfsonar

domains, allowing engineers to analyse and diagnose network behaviours across the entire end-to-end path. The tools provided in the perfSONAR suite perform active measurements of throughput, packet loss, delays and jitter, and record network route and path changes. Some features include:
    - uniform interface that allows for the scheduling of measurements, storage of data in uniform formats, and scalable methods to retrieve data and generate visualizations.
    - extensible system to support new metrics
    - multiple possibilities for data presentation

As mentioned above, RELIANCE is and will be making use of monitoring services in order to be able to verify the compliance of its services to any intended SLA. Integration with a monitoring infrastructure from EOSC may be reconsidered at later stages of the project, particularly with EGI service monitoring as described in the documentation[72], taking into account the evolution of the relevant policy and tools in EOSC.

7.2.3 *Accounting services*

The RELIANCE services will consume storage and compute capacity from underlying provider(s). Currently, PSNC is playing this infrastructure provider role, supporting the storage and compute needs of RELIANCE services. However, in order to support a larger number of users and communities, as targeted, the underlying infrastructure will need to scale out and grow horizontally. In order to achieve this, RELIANCE plans to leverage the EOSC resources like storage and cloud compute (as described below). Accordingly, having a full understanding of the use of resources (e.g., per community or per user basis) would help RELIANCE to build the picture of its operational costs and plan sustainable business model for beyond the project lifetime.

As introduced in Section 5.3.3, the EOSC Accounting system is based on the EGI Accounting system:

- The EGI Accounting service[73] stores user accounting records from various services. It works thanks to a network of message brokers that transfer usage data from the host to a central repository of information. The data is handled securely and can be consulted online through the Accounting Portal[74]. Accounting gives:
    - Increased control over resource consumption
    - Reduced overhead of defining data models, architecture and setup of an accounting system
    - Reduced cost of maintaining an accounting infrastructure
    - Access to a reliable, high available, high performance service
    - User friendly web interface

Accounting the utilization of physical (or virtualized) resources, though, is only one of the accounting aspects towards defining a sustainable model. RELIANCE services will also need to take into account aspects like API calls (per x time), number of resources, logins, etc., which target different project KPIs. Therefore, RELIANCE will carry out an internal accounting of those aspects, and although not part of the main plan, it will analyse integration with EOSC accounting[75] for the future.

[72] https://confluence.egi.eu/display/EOSC/Monitoring+Service
[73] https://www.egi.eu/internal-services/accounting/
[74] https://accounting.egi.eu/
[75] https://confluence.egi.eu/display/EOSC/Accounting

7.2.4 *Helpdesk*

As introduced in Section 5.3.4, the EOSC helpdesk system is a ticketing service that connects EOSC users and service providers. RELIANCE could leverage this system to respond to user requests and any other type of issue in the EOSC context. EOSC portal offers two helpdesk services:

- EGI Helpdesk[76] provides information and support to troubleshoot problems with products and services. It is the central contact point of the EGI e-Infrastructure federation, where users can request user support, report incidents and problems, submit service requests, or raise new requirements to EGI. The system is a distributed helpdesk with central coordination. The main characteristics include:
    - Central point of contact for support
    - Repository of information and solutions
    - Keeps track of progress of ongoing issues happening on the infrastructure
- Open Science Helpdesk[77] supports stakeholders' questions on Open Science and it builds a knowledge base on a range of topics targeted at different stakeholders. It includes a 24x7 Helpdesk supported by NOADs, FAQs, Factsheets, Briefing Papers and topical webinars. It provides direct and round the clock access to OpenAIRE support for wide ranging issues on Open Science across European Research Institutions, Projects and Individuals.

For RELIANCE, it is particularly interesting the EGI Helpdesk, which can be easily reused/integrated in different EOSC services. Service providers can choose from different integration ways, and then open a ticket providing the service information and selected integration, as described in Section 5.3.4. In RELIANCE, although there are no specific requirements on the provisioning of a helpdesk service, the plan is to offer such a system based on the resources and experience of PSNC with the support of the technical coordinator (Terradue). Integrating this service with EGI Helpdesk would be considered in a later stage of the project.

7.3 Research data services

In this section we provide an overview of some of the existing EOSC services dealing with FAIR data management that we plan to reuse and connect with the RELIANCE services. As the project evolves and as the EOSC service portfolio expands we may consider additional services for integration.

In this plan we are already taking into account the planned services of the ongoing sister projects of INFRAEOSC-07 call, including DICE, EGI-ACE, OpenAIRE-Nexus and C-Scale. However, as these are being implemented at the same time as RELIANCE, we need to be flexible and follow closely their developments in order to adjust/update our plan accordingly. We have to bear in mind that the EOSC landscape is evolving rapidly and that, in fact, we are contributing to its evolution.

As described in Section 5.4, the EOSC offers various services for service providers to help them to manage research data and implement scenarios where such data needs to be stored, transferred, analysed, indexed, etc. Taking as reference the virtuous cycle of research, we list below the relevant categories and the corresponding services that we are planning to reuse.

- Access, deposition and sharing
    - Zenodo
    - B2SHARE
    - B2Drop
- Processing and analysis
    - Cloud compute, e.g., federated IaaS & PaaS (potentially)

[76] https://helpdesk.egi.eu/

[77] https://www.openaire.eu/support/helpdesk

    - Jupyter Notebook
- Discover and reuse
    - OpenAIRE Research Graph (potentially)
- Data Management, curation & preservation
    - Argos (potentially)

It is important to note that these services are part of the offering of the above mentioned projects. In particular, B2SHARE and B2DROP are part of the DICE service portfolio, as illustrated in the Figure 10 below with the research data workflow by DICE.

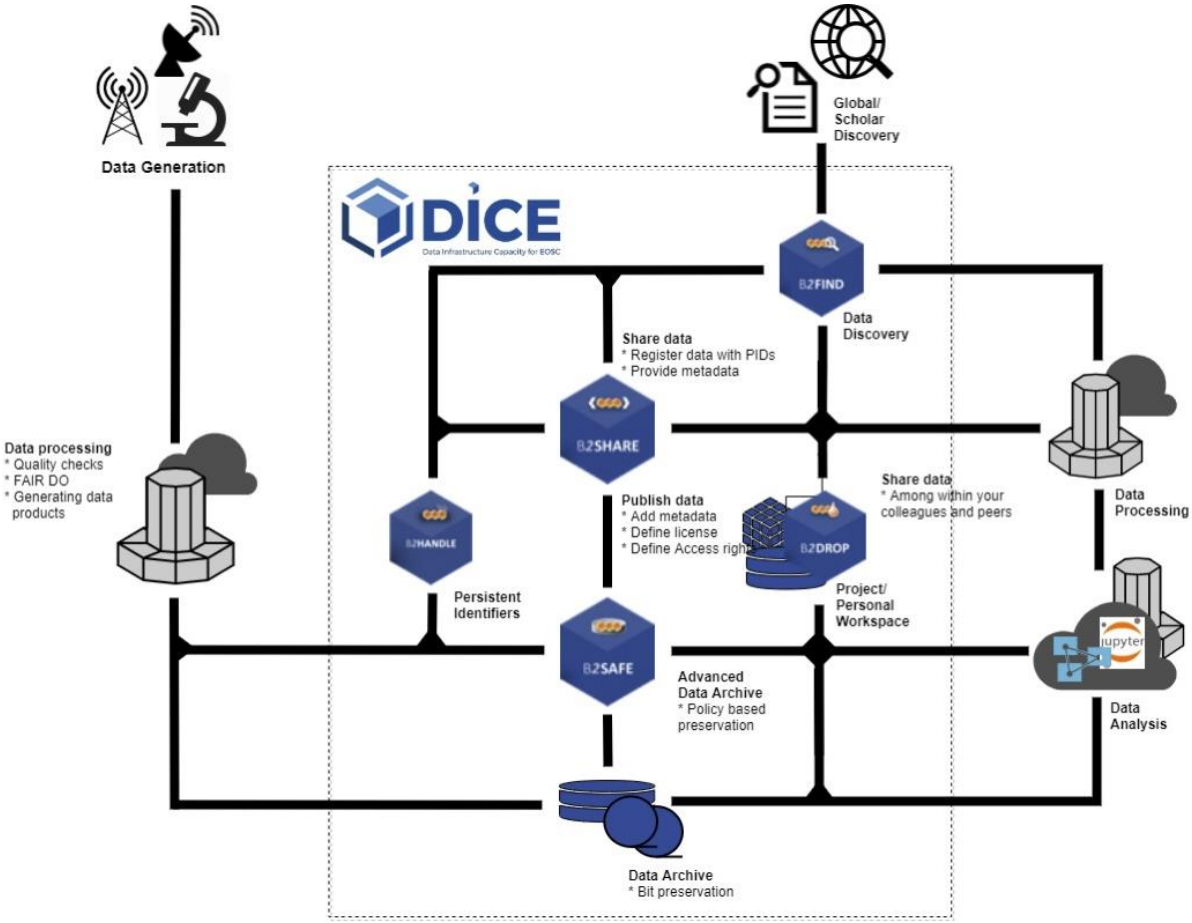

Figure 10: DICE Research Data Workflow

Similarly, the processing and analysis services are part of the EGI-ACE service portfolio, as illustrated in the project concept diagram shown below (Figure 11). In the diagram we can highlight the Interactive Computing services that include platforms like Jupyter Notebooks, and the federated access and resources of cloud computing enabling IaaS, PaaS and more. Also, the diagram shows that EGI-ACE includes the EOSC AAI (EGI check-in) in their offering (Federated Identity).

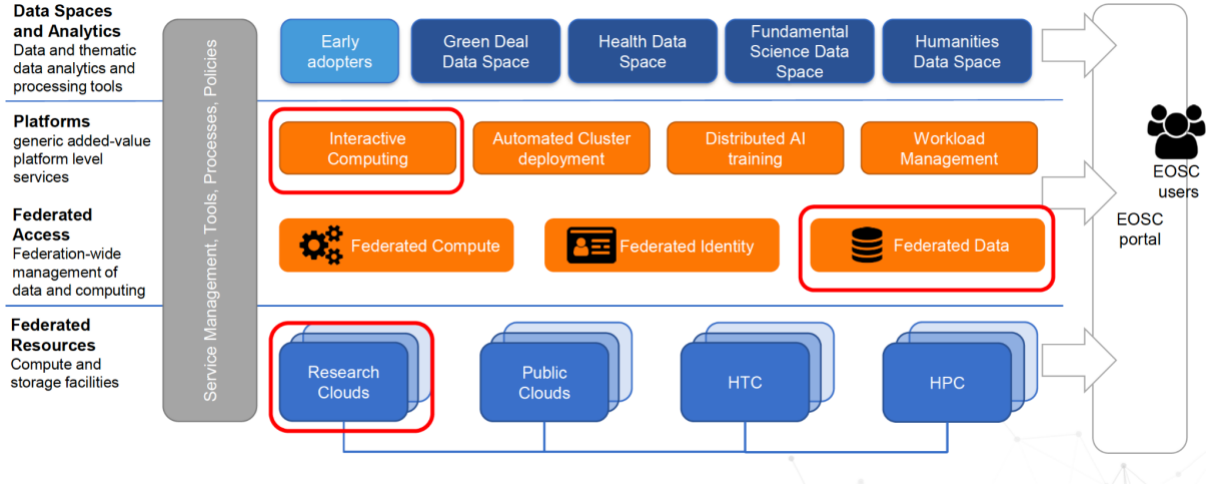

Figure 11 EGI-ACE concept and methodology

Finally, services like Zenodo, OpenAIRE Research Graph and Argos are part of the OpenAIRE-Nexus service portfolio as illustrated in the Figure 12 below.

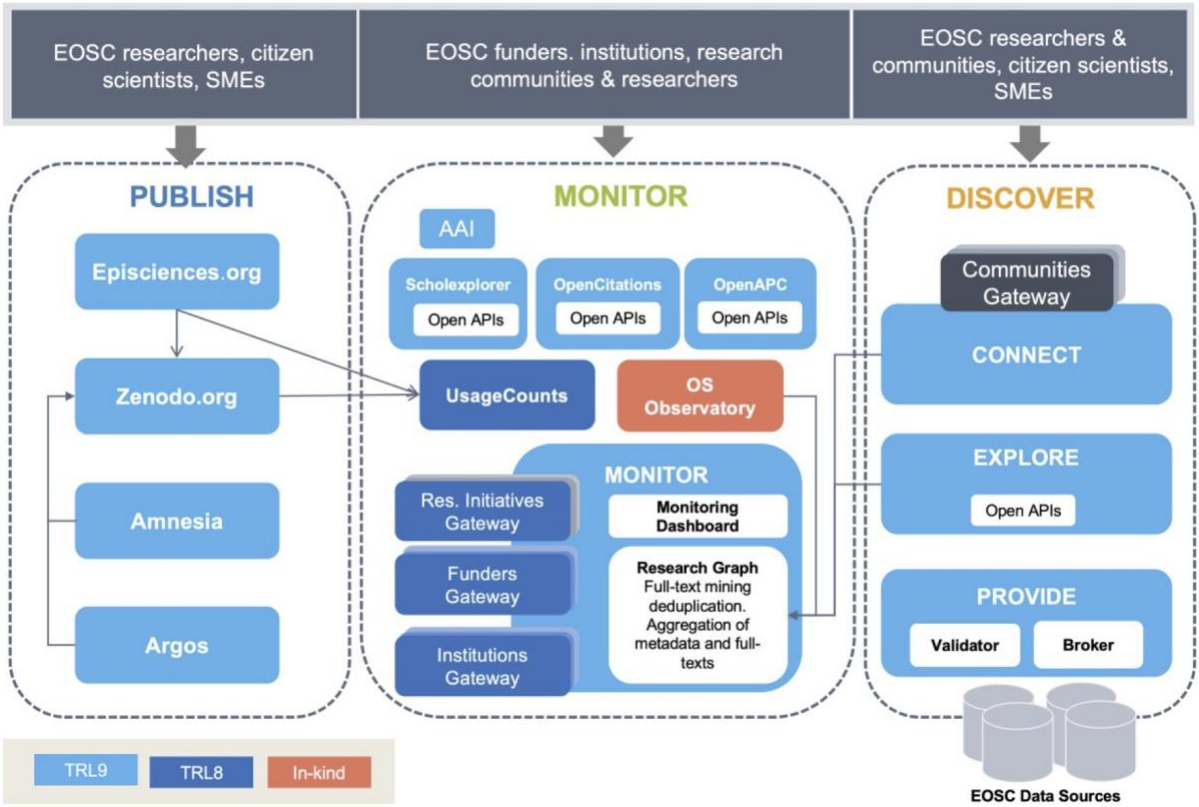

Figure 12: OpenAIRE-NEXUS service portfolio

In the following we briefly discuss the selected services and how they will be used by RELIANCE.

7.3.1 *B2DROP*

B2DROP[78] is a secure and trusted data exchange service for researchers and scientists to keep their research data synchronized and up-to-date and to exchange with other researchers. B2DROP allows to store and exchange data with colleagues and team members, synchronize multiple versions of data, and ensure automatic desktop synchronization of large files. It is particularly useful for researchers that need to synchronize and exchange data with one or multiple users. Main features include:

- Web access
- Sharing a file to local users
- Share publicly across B2DROP instances
- Accessing shared artefacts
- Desktop synchronization
- Mount B2DROP using the WebDAV client
- Publishing files to B2SHARE
- Default quota of 20GB

RELIANCE aims to leverage B2DROP to provide a place where researchers can use as their personal workspace to store (intermediate) results and data that is used/produced during the course of their investigations.

In particular, as the resources aggregated by a research object in ROHub can be internal and/or external to the system, provided that they are accessible via some URL, ROHub typically operates alongside personal storage services where scientists can easily save the auxiliary resources produced/consumed during the research lifecycle. Similarly, scientists may need a place to store (temporal) data produced by ADAM service to access EO datasets. Accordingly, RELIANCE aims to leverage this kind of storage facilities from EOSC offering, it order to support a larger number of scientists and to facilitate them a place where they can keep the resources aggregated by their research objects. Ideally, such integration would be transparent for the users. This, however, requires further analysis on the possible ways to integrate B2DROP in ROHub, which is an on-going task.

7.3.2 *B2SHARE*

B2SHARE[79] is a solution for researchers, scientific communities and citizen scientists to store and publish small-scale research data from diverse contexts. It facilitates research data storage, guarantees long-term persistence of data and allows data, results or ideas to be shared worldwide. It is particularly useful for researchers that do not have adequate facilities for storing, preserving and sharing data, providing them with a safe repository for scientific data and an easy way to share it in the research community. The basic production service comprises the following features:

- self-service registration for any scientists and researchers,
- free upload and registration of stable research data,
- data access policy is defined by the data owner,
- metadata is openly accessible and harvestable,
- customized metadata handling and customized user interfaces (e.g., for metadata acquisition),
- data integrity is ensured by checksums which are calculated during data ingest,
- the data is kept online, the storage usage is based on the principle of fair share

B2HARE offers a REST API that can be used for interaction with B2SHARE via external services or applications, for example for integration with other web-sites (research community portals) or for

[78] https://b2drop.eudat.eu/
[79] https://b2share.eudat.eu/

uploading or downloading large data sets that are not easily handled via a web browser. The API can also be used for metadata harvesting, although an OAI-PMH API endpoint is also provided for this purpose.

RELIANCE's plan to reuse B2SHARE is twofold. On the one hand, it will be used as a source of research data/resources. Research objects connect related scientific resources (from the same investigation, observation, research work, etc.) including, among others, results, articles or other forms of scholarly communication that may be deposited in B2SHARE. In this sense, RELIANCE will enable the connection of these resources with other related resources (e.g., data, methods) available in different repositories or locations via research objects, along with a rich set of metadata (provenance, relations, evolution, etc.). On the other hand, as part of the research object evolution, different snapshots/releases can be generated throughout the research lifecycle. These snapshots/releases can be then published in research repositories or scholarly communication platforms, such as B2SHARE and Zenodo for their long-term preservation and citation. The plan is to give the possibility to the researchers to publish those releases into B2SHARE directly from ROHub in a transparent way.

7.3.3 *Zenodo*

Zenodo[80] is a catch-all general purpose repository developed under the OpenAIRE program and operated by CERN, in support of Open Science and FAIR principles. It enables researchers, scientists, projects and institutions to share, preserve and showcase research results (data, software and publications) that are not part of the existing institutional or subject-based repositories of the research communities. It is currently used by more than 50K researchers and 3K communities all over the world. The main features include:

- It accepts all research outputs from across all fields of research and any file format including both positive and negative results.
- It assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citable
- It allows you to create collections (communities) and to accept or reject uploads submitted to it.
- It stores data safely for the future in the same cloud infrastructure as the research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software Invenio
- It is integrated into reporting lines for research funded by the European Commission via OpenAIRE
- It encourages to share research as openly as possible to maximize use and re-use; however, it allows for uploading under a variety of different licenses and access levels.

Zenodo supports harvesting of all content via the OAI-PMH protocol, and it offers a REST API to allow communication with other services. The current API supports depositing (upload and publishing of research outputs), searching for published records and to download/upload files.

RELIANCE's plan to reuse Zenodo is quite similar as for B2SHARE. It will be used as a source of research data/resources. RELIANCE will enable the connection of Zenodo resources with other related resources (e.g., data, methods) available in different repositories or locations via research objects, along with a rich set of metadata (provenance, relations, evolution, etc.). On the other hand, RELIANCE plans to give the possibility to the researchers to publish research objects snapshots/releases into Zenodo directly from ROHub in a transparent way.

The key role that research objects can play in the implementation of workflows for continuous publishing of research products in scholarly communication services has been already highlighted by

[80] http://zenodo.org/

Zenodo/OpenAire[81] [9], which proposes the use of research objects as a further step to enhance the publishing workflows and to allow monitoring of the impact of Research Infrastructures (RI) used for research. The idea is to publish a research package, including not only a dataset but also metadata information, and references to the used RIs, and then assign this package a DOI to make it a citable object. RELIANCE will help to realize and enhance such vision by enabling the publishing of research objects from ROHub into Zenodo.

7.3.4 *Cloud compute*

EGI Cloud Compute[82] allows the deployment and scaling of virtual machines on-demand. It offers guaranteed computational resources in a secure and isolated environment with standard API access, without the overhead of managing physical servers. Cloud Compute offers the possibility to select pre-configured virtual appliances (e.g., CPU, memory, disk, operating system or software) from a catalogue replicated across all EGI cloud providers. The main features include:

- Execute compute- and data-intensive workloads (both batch and interactive).
- Host long-running services (e.g., web servers, databases or applications servers).
- Create disposable testing and development environments on virtual machines and scale the infrastructure needs.
- Select virtual machine configurations (CPU, memory, disk) and application environments to fit specific requirements.
- Manage the Cloud Compute resources in a flexible way with integrated monitoring and accounting capabilities.

RELIANCE's plan to leverage EGI Cloud compute is also twofold. First, RELIANCE may need to rely on EGI Cloud Compute resources to enable its services to scale-out and grow horizontally. For instance, ADAM may need dedicated storage and high computing capabilities to handle larger amounts of EO data. The Copernicus EO data pipelines may also need to leverage and exploit resources offered by EGI Cloud Compute to create and deploy different pipelines in order to support a large number of users and research communities. Additionally, RELIANCE plans to leverage EGI Cloud Compute resources for its users and research communities, who typically require computing resources (IaaS, PaaS) to run their community-specific tools/applications where they carry out their research work. For example, two of our research communities are already using VMs provided by PSNC to carry out their research activities. However, in order to support a much larger number of users and research communities, as it's our target, we plan to rely on EGI Cloud Compute resources.

7.3.5 *Jupyter Notebooks*

EGI Notebooks[83] is a browser-based tool, based on JupyterHub technology, for interactive analysis of data using EGI storage and compute services. This service can combine text, mathematics, computations and their rich media output using Jupyter technology, and can scale to multiple servers and users with the Cloud Compute service.

Notebooks for Researchers: After a lightweight approval, users can login, write and play notebooks using storage and compute capacity. For individual users, notebooks enable reproducible research (notebooks can be re-played by the same user, shared and re-played by different users), and they can easily hook into other big-data environments (e.g., Spark, Hadoop) or services (e.g., Cloud Compute) provided by or hosted by EGI.

[81] https://zenodo.org/record/3701394#.YGMrgGQzajA
[82] https://marketplace.eosc-portal.eu/services/egi-cloud-compute
[83] https://marketplace.eosc-portal.eu/services/egi-notebooks

Notebooks for Communities: EGI offers consultancy and technology to set up a community-specific JupyterHub on top of a community VO. Comes together with EGI-enabled compute and storage resources and with community-specific storage.

RELIANCE aims to exploit Jupyter notebooks, containing code, equations and other interactive content, as the processing and analysis environment for research objects. Accordingly, RELIANCE plans to leverage the EGI Notebook facilities for interactive computing. The goal is to enable scientists to search for, open, and exploit the research object content and metadata from the notebooks using the provided python library, and to save the results and provenance information back to the research object.

7.3.6 *OpenAIRE Research Graph*

The OpenAIRE Research Graph enables developers to realize services for scholarly communication and research analytics. It exposes an API that allows the access to the OpenAIRE Graph, a scholarly communication graph, including information about objects of the scholarly communication life-cycle (publications, research data, research software, projects, organizations, etc.) and semantic links among them. The service gives access to the OpenAIRE Graph via different protocols (OAI-PMH, HTTP API, SPARQL) to serve developers with different requirements and preferences. The OpenAIRE Graph is created bi-monthly by:

- Aggregating metadata from OpenAIRE's European and global network of validated data sources;
- Enriching metadata by full-text mining and inference: inference results (metadata records and relationships) are included in the graph in the form of a data source;
- Collecting metadata from end-users via the EXPLORE service: users' feedback (metadata records and relationships) is included in the graph in the form of a data source.

RELIANCE is evaluating the possibility to leverage the OpenAIRE Research Graph, in particular in the Text Mining services for the implementation of the search engine and the recommendation system to use it as a source of research resources.

7.3.7 *Argos Data Management Plan (DMP) tool*

ARGOS[84] is an online tool that supports the automatic processes of creating, managing, sharing and linking DMPs with associated research artifacts. The tool delivers an open platform for data management planning that addresses FAIR and Open Data best practices and assumes no barriers for its use and adoption. It applies common standards for machine-actionable DMPs as defined by the global research data community of RDA and incorporates feedback received from researchers, research communities and funders. The tool allows actors (researchers, managers, supervisors, etc.) to create actionable DMPs that may be exchanged among infrastructures to carry out specific aspects of the data management process in accordance with the intentions and commitment of data owners.

RELIANCE plans to deliver its initial DMP at M6, and then, an updated version at the end of the project. We are currently evaluating using Argos to support this task, in order to facilitate the creation, maintenance of the plan and to document it in a more standard format.

[84] https://argos.openaire.eu/home

8 Integration plan roadmap

In this section we describe the plan for integrating RELIANCE services into the EOSC. Our key goals are:

- Register and onboard the core RELIANCE services in EOSC, including ROHub, ADAM, Text Mining, as well as the pipelines for Copernicus data, so that they will be available via the EOSC Portal Catalogue and Marketplace
- Integrate RELIANCE services with relevant EOSC services that provide added-value to our offering (e.g., to enhance our operations, to deliver more features to users)
- Develop the RELIANCE IT Service Management System gradually, and in an EOSC compatible way.

Similar to the approach proposed by NEANIAS project, we aim to achieve these goals following an agile approach, although based on three compulsory and one optional successive development cycles, considering the following:

- The cycles will gradually increase the degree of integration of RELIANCE services, from a basic EOSC integration to more complex EOSC integration with advanced capabilities.
- The cycles are dynamically defined. We start with a clear definition of the first two cycles and with a definition of the goals for the next ones; however, these other cycles will be confirmed/updated after reaching the initial ones, taking into account the evolution of the relevant EOSC services, i.e., the third cycle will be updated after reaching the second one.
- The cycles are applicable for each individual service; however not necessarily every service will follow every cycle. In particular, cycle one and two are applicable for all the services, but cycles 3 and 4 are applicable only to some of them.
- Each service may finalize a cycle at a different point in time, e.g., one service may finalize a cycle before another; however, the roadmap will define a target date for the core services, when that cycle should be finalized (when applicable), while additional services (e.g., pipelines for Copernicus data) may carry out these cycles at later stages when they will be more mature.
- The cycle one will start as soon as it is realistically possible, so that RELIANCE services can be used by researchers from an early stage
- The cycle one should be finalized before the open call is launched, so that new communities can see and try the services before applying

The cycles are as follow:

- **Cycle 1 - Services registration and onboarding in EOSC**: Following the guidelines of the EOSC onboarding process, we will:
    - Complete the Join as a provider form on the EOSC Portal website
    - Fill-in the service Description Template (SDT)
    - Ensure the services meet the “Rules of Participation” specified for the EOSC portal
    - Review and validate the service entry in the portal
- **Cycle 2 - Integration with EOSC AAI**: We will enable the authentication to RELIANCE services via the EOSC AAI services. In particular we plan to integrate with EGI check-in and/or eduTEAMS, following the services provider integration workflow. In EGI case, for example, we will[85]
    - Register our Service Provider and test integration with the demo instance of EGI Check-in.
    - Register our Service Provider with the production instance of EGI Check-in to allow members of the EGI User Community to access the service.
    - Ensure service meets eligibility criteria

[85] https://docs.egi.eu/providers/check-in/sp/

- **Cycle 3 - Integration with EOSC Research Data Services**: We will integrate the applicable RELIANCE services with the selected EOSC Research Data services, including B2DROP, B2SHARE and ZENODO, we will reuse Cloud Compute and Notebooks services as needed and potentially the OpenAIRE Research Graph
    - B2DROP service will be used (and potentially integrated) by ROHub to provide researchers with a scalable and secure personal workspace to store their resources. If possible, the service will be integrated in ROHub transparently, allowing users to specify if they want to use B2DROP as the default storage for their resources in ROHub. B2DROP may also be used by ADAM to store working data cubes generated by the researchers to access EO data.
    - B2SHARE will be reused and integrated in ROHub, allowing users to specify if they want to deposit and publish snapshots/releases of their research objects in this repository. Potentially, users will be also to select the appropriate community where to publish the research object.
    - Zenodo will be reused and integrated in ROHub, allowing users to specify if they want to deposit and publish snapshots/releases of their research objects in this repository. Potentially, users will be also to select the appropriate community where to publish the research object.
    - Cloud compute resources will be used as needed both by service providers and by user communities. For services providers, they may use those resources to scale out their services, while user communities may use those resources to run their community applications/tools to carry out their research work.
    - Notebooks service will be used as needed by RELIANCE research communities in order to run processing and analysis tasks. RELIANCE will provide python libraries to enable researchers to use and exploit research objects and data cubes within the notebooks.
    - OpenAIRE Research Graph will be potentially used by the Text Mining services for the implementation of the search engine and the recommendation system to use it as a source of research resources.
- **Cycle 4 - Integration with EOSC Support Services (optional)**: RELIANCE is or will use appropriate supporting services like monitoring services to verify compliance to any intended Services SLA, internal accounting of resources, and helpdesk services. However, Integration with a EOSC support federation services may be considered at later stages of the project.

Table 1 below summarizes the cycles of the integration roadmap, mapping the EOSC services with the applicable RELIANCE services, and the target date for finalizing the cycle. As discussed above the cycles are dynamic and may be adjusted/updated after finishing a previous cycle, taking into account the evolution of the EOSC landscape, and in particular results from other INFRAESOC-07 projects.

Table 1: RELIANCE integration plan roadmap (are optional services)*

cycle	EOSC service category	EOSC service	RELIANCE services	Max date
1	EOSC portal	catalogue and marketplace	ROHub ADAM Text Mining EO data pipelines (may be later)	M6
2	EOSC AAI	EGI check-in (& eduTEAMS)	RELIANCE Identity and Access Management (Keycloak) -	M9

			used by all services with authentication, e.g., ROHub, ADAM, Text Mining	
3	EOSC Research Data services	B2DROP B2SHARE Zenodo Cloud Compute Notebooks OpenAIRE Research Graph* Argos*	ROHub ADAM Text Mining Copernicus EO data pipelines	M14
4	EOSC support services*	EGI service monitoring EOSC accounting EGI Helpdesk	RELIANCE monitoring services (sentry, Nagios) RELIANCE accounting services RELIANCE helpdesk	M24*

9 Conclusions

RELIANCE aims to deliver a suite of innovative and interconnected services that extend EOSC's capabilities to support the management of the research lifecycle within Earth Science Communities and Copernicus Users. The RELIANCE service portfolio is based on key and complementary technologies including:

- ROHub platform providing Research Object management functionalities
- ADAM platform providing Data Cubes management functionalities
- Text Mining services providing functionalities for information extraction and semantic annotation, content-based retrieval and recommendation, as well as extended analytics services in support of the scientific enterprises
- Copernicus Data Pipelines, an added-value service enabling the systematic execution of EO applications to continuously deliver data and information to different users, complying with their specifications for information content and format.

In this deliverable, we discussed the EOSC integration guidelines for service providers, which specify the registration and onboarding of the services into the EOSC portal as a minimum requirement, the federated support services offered by EOSC and an overview of the available research data services. Next, this document provides a detailed description of RELIANCE services, highlighting how they plan to leverage selected EOSC services to bring added value to their users and research communities.

In particular, based on the current EOSC offering we defined our integration plan as follows:

- We will register and onboard the RELIANCE services in the EOSC Portal. This will include completing the Join as a provider form on the EOSC Portal website, fill-in the service Description Template (SDT) and ensure the services meet the "Rules of Participation" specified for the EOSC portal.
- We will integrate RELIANCE services with the EOSC AAI service. This will include registering our RELIANCE AAI service and test the integration with the demo instance of EGI Check-in, and then registering it with the production instance, ensuring the compliance with the eligibility criteria.
- We will integrate the applicable RELIANCE services with the selected EOSC Research Data services, including B2DROP, B2SHARE and ZENODO, we will reuse Cloud Compute and Notebooks services as needed and potentially the OpenAIRE Research Graph
- Finally, we will potentially leverage and integrate with the federated EOSC Support Services. For this, we will finalize the deployment of all the RELIANCE IT service management processes to handle the monitoring, accounting and helpdesk, aligning with the corresponding EOSC-hub SMS processes, and potentially defining the RELIANCE Data Management Plan with the support of the available services.

We aim to achieve these goals following an agile approach based on three compulsory and one optional successive development cycles, where the cycles are dynamically defined. So, after finalizing the two cycles, the next cycles will be confirmed/updated, taking into account the evolution of the relevant EOSC services.

Moreover, we will be closing monitoring and cooperating with other INFRAESOC-07 projects, to follow their latest advances. We will also monitor other EOSC related projects and working groups including the upcoming EOSC future project, and the working groups on architecture and on interoperability. Therefore, we will stay open for new integration opportunities offered by EOSC, and reconsider/update our integration roadmap as needed. The integration plan and its implementation will be assessed, and revised if needed, in the Annual report on EOSC operation in M12 of the project, and the final results will be presented in the Annual report on EOSC operation in M24 of the project.

References

[1] Sipos, Gergely. (2020, May 14). EOSC-hub Integration handbook for service providers (Version 14/May/2020). Zenodo. http://doi.org/10.5281/zenodo.3826907
[2] NEANIAS Deliverable D8.1, EOSC integration plan, Nov 2019
[3] European Commission. Directorate-General for Research, & Innovation (Feb, 2021). EOSC Architecture Working Group View on the Minimum Viable EOSC. Report from the EOSC Executive Board Working Group (WG) Architecture. Publications Office of the European Union. DOI: 10.2777/492370
[4] European Commission. Directorate-General for Research, & Innovation (Feb, 2021). EOSC interoperability framework. Report from the EOSC Executive Board Working Groups (WG) FAIR and Architecture. Publications Office of the European Union. DOI: 10.2777/620649
[5] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18
[6] European Commission. Directorate-General for Research, & Innovation (Nov, 2020). Solutions for a Sustainable EOSC: A FAIR Lady (olim Iron Lady) Report from the EOSC Sustainability Working Group. Publications Office of the European Union. DOI: 10.2777/870770
[7] EOSCpilot Deliverable: D5.4: Final EOSC Service Architecture, Apr 2019
[8] Palma R., Corcho O., Gómez-Pérez J.M., Mazurek, C. ROHub - A Digital Library of Research Objects Supporting Scientists Towards Reproducible Science. In Semantic Publishing Challenge of Proc. Extended Semantic Web Conference (ESWC), Crete, Greece, May 25-29, 2014."
[9] Bardi, Alessia, Manunta, Michele, Toth-Czifra, Erzsebet, Vergoulis, Thanasis, Manghi, Paolo, & Baglioni, Miriam. (2019). D7.3 – Interoperability with Research Infrastructures. Zenodo