JERICO-DS Deliverable Joint European Research Infrastructure network of Coastal Observatories **Design Study**	
WP, Deliverable and full title	JERICO DS – D8 / D3.2 - WP3 - e-JERICO Technical Design Study
5 Key words	JERICO-RI, JERICO-CORE, J-CORE, Federated Cloud-Computing Infrastructure, Architecture of Services.
Lead beneficiary	SOCIB
Lead Authors	Juan Miguel Villoria (SOCIB)
Co-authors	Miguel Ángel Alcalde (SOCIB), Miguel Charcos Llorens (SOCIB), Juan Gabriel Fernández (SOCIB), Emma Reyes (SOCIB), Joaquín Tintoré (SOCIB)
Contributors	
Submission date	2023.06.01
Nature	Report

GRANT N°: 951799
PROJECT ACRONYME : JERICO-DS
PROJECT NAME : Joint European Research Infrastructure of Coastal Observatories – Design Study
COORDINATOR : Laurent DELAUNEY - Ifremer, France - jerico@ifremer.fr

Document information	
Document Name	*JERICO-DS_D8/D3.2*
Document ID	*JERICO-DS-WP3-D8/D3.2-2023.06.01-V2.0*

JERICO-DS Milestone
Joint European Research Infrastructure network of Coastal Observatories
Design Study

Revision	Date	Modification	Author
V0.1	2022.07.0	First Version	Juan Miguel Villoria
V0.2	2022.09.1	Introduction. Analysis of State of the Art. Strategic Collaborations. Evolution of Concept. Requirements.	Miguel Charcos Llorens
V0.3	2022.12.0	JERICO-CORE Service Architecture. Appendix A - Requirements.	Juan Miguel Villoria Miguel Ángel Alcalde Miguel Charcos Llorens Juan Gabriel Fernández
V0.4	2023.02.0	Cloud Computing Model.	Juan Miguel Villoria
V0.5	2023.05.11	Implementation Solutions. - EGI - EUDAT - EOSC - D4Science - Blue-Cloud.	Juan Miguel Villoria
V0.6	2023.05.	Evaluation.	Juan Miguel Villoria
V0.7	2023.05.2	Conclusions.	Juan Miguel Villoria
V0.8	2023.05.3	JERICO-DS Technical Design Study in the ESFRI Roadmap.	Juan Miguel Villoria Juan Gabriel Fernández
V1.0	2023.06.0	Final review.	Miguel Ángel Alcalde Miguel Charcos Llorens Juan Gabriel Fernández Emma Reyes Joaquín Tintoré

Diffusion list			
Consortium beneficiaries	Third parties	Associated Partners	Others
X	X		

TABLE OF CONTENT

1. EXECUTIVE SUMMARY

This deliverable presents the outcome of JERICO-DS Task 3.3: e-JERICO (from now on JERICO-CORE) Technical Design Study.

The JERICO-CORE Technical Design Study aims to envision the federated cloud-computing infrastructure needed to serve all users (researchers, governmental agencies, private companies and general public), promoting the use of data, information, and services for academic, industrial, and societal purposes offered by JERICO-RI.

The JERICO-RI will be the European gateway to long-term scientific observations and related services for European coastal marine systems at the convergence between the land, open ocean, and atmosphere; empowering European research excellence and expertise for the benefit of society.

The JERICO-DS Technical Design Study, establishes the foundations and guidelines to the future develop of JERICO-CORE platform:

- To define a Service Architecture based on well-gathered and analysed stakeholder requirements.

- To embrace the latest technologies in cloud computing, suitable for use as tools for analysis, processing, and communication of the scientific production of JERICO-RI.

- To not reinvent the wheel, and develop JERICO-CORE on the major existing European Initiatives (EGI, EOSC, EUDAT, D4Science and Blue-Cloud), in order to reduce risks in development, quickly gain knowledge, experience and return on investment in terms of costs and time to release.

2. INTRODUCTION

The coastal ocean environments are important for our society because many communities depend to a large degree on oceans, seas, and coastal biodiversity for their livelihoods. They provide an inestimable source of economic, social and cultural wealth. However, they are largely influenced by climate changes and other local stressors such as pollution and overfishing. Understanding the impact of these global and local pressures requires managing and assessing observational marine and coastal data, models and other resources involved in the creation, analysis and management of reliable data. In the last few years, a significant number of e-infrastructures and information systems have been developed for this purpose. They led to the creation of a significant amount of data, information and knowledge that is scattered among different Research Infrastructures (RI) and knowledge holders. Therefore, there is a growing need to standardise the information and create common best practices that allow for the necessary level of coordination between these infrastructures and lead to the appropriate access and use of information for a sustainable development of the ocean. In this context, the UN Decade of Ocean Science for Sustainable Development, 2021–2030, the 'Ocean Decade', was put in place to bring together ocean actors worldwide to promote the partnership needed to generate and integrate ocean knowledge. The Ocean Decade established, under the Secretariat of the Intergovernmental Oceanographic Commission (IOC), the Ocean Decade Data Coordination Group committed to help generating and sharing large amounts and types of data, information and knowledge that support the goals of the Ocean Decade. Their ambition is to create a global digital ecosystem that addresses coordination between data generators and users from cross-sectorial stakeholders including governments, research institutions, industry and public. In this study of the technical design of the JERICO-CORE platform, we will establish the link with broader digital ecosystems. This link will allow both the reuse and alignment with the underlying technologies, systems and best practices, and the interoperability with such as digital systems.

The creation of a global ecosystem is particularly important because the use of digital twins in the ocean field is rapidly expanding worldwide to serve a wide range of use cases, from policy-makers to marine environment management, science, and operational services. The Digital Twins of the Ocean program (DITTO) was endorsed by the Ocean Decade to coordinate co-design of Digital Twins of the Ocean (DTO). Its mission is to promote and define common best practices for professionals of all sectors to create custom DTOs. It will be an essential framework to predict climate change, extreme events, and to evaluate scenarios derived from political decisions or changes in social habits. In this context, big data, artificial intelligence and other advanced data technologies will be key to the development of a DTO ecosystem that allows the creation of virtual collaborative environments and facilitates stakeholders to build upon knowledge, tools, services and best practices. In these environments, analytics of real time big data will require trustable, interoperable and standardised information that allows scientific models to use the state of the art science and knowledge to produce the desired outcomes.

The global approach of the Ocean Decade follows the already existing trend of the European Commission (EC) to find common ways to improve data flows, enable interoperable data sharing and stewardship, and enhanced collaboration between data providers and users. For this purpose, the EC funded the implementation and development of the ENVironmental Research Infrastructures (ENVRI) community that brings together projects, networks, RIs and e-infrastructures supporting their data solutions. ENVRI (https://envri.eu/) facilitates interoperability of knowledge between members and other stakeholders by harmonising products, services and best practices in a way that promotes synergies towards a joint vision and strategy to streamline the RIs activities. ENVRI supports maturing national and regional infrastructures with the know-how of associated RIs at the level of the European Strategy Forum on Research Infrastructure (ESFRI). The ESFRI (https://www.esfri.eu/) forum supports a coordinated European approach between policy makers and research, leading to addressing the challenges of the UN Sustainable Development Goals. It ensures that RIs are openly and easily available across Europe and promotes the interoperability of European e-infrastructures, information, data and knowledge. The ENVRI community advocates interoperability between infrastructure, cross-disciplinary collaboration, discoverability and accessibility of dispersed data and information across many different sources and in many different formats. During the ENVRIPlus project it was identified the need to achieve harmonisation of data and services as well as their consistent representation, interpretation and access, both within and between RIs of ENVRI. Within this approach, ENVRI has participated in the European Grid Initiative (EGI) and the European Open Science Cloud (EOSC) initiatives that support the interoperability of RI information across fields. In the later ENVRI-FAIR project, the ENVRI community advanced interoperability in particular but also FAIRness in general of the data and services offered by the ENVRI members. These developments will be critical to the integration of knowledge into EOSC.

The marine ecosystem presents specific challenges due to the diversity of observed conditions from coastal areas out to deep sea, with a large variety of depths and environments between oceans' surface to the seafloor. In order to face the complexity of the ocean conditions, the RI's landscape for ocean research has become increasingly complex with a large diversity of data infrastructures. The EC has focused on integrating European ocean data infrastructures and setting up programs to coordinate initiatives related to ocean research and management that supports the ambition of the Ocean Decade of creating a DTO. European DTOs will facilitate the discoverability of ocean knowledge to citizens, entrepreneurs, scientists and policy-makers in order to face the challenges of marine and coastal habitats. The European DTO will build on existing infrastructures, in particular the large data providers including SeaDataNet, EMODnet, Copernicus Marine Service and Data Information Access Services (DIAS). However, building such a digital system that transforms data into knowledge and information requires a deep understanding of the assets that helped to produce reliable data feeding DTO models.

The Joint European Research Infrastructure network for Coastal Observatories (JERICO) is a critical piece to allow models of the DTO to connect continental, atmospheric and open ocean RIs, thus filling a key gap in the ESFRI landscape. It is an integrated pan-European multidisciplinary and multi-platform research infrastructure dedicated to a holistic appraisal of coastal marine system changes. JERICO-RI (https://www.jerico-ri.eu/) aims to provide an integrated solution in Europe to face the challenges of the coastal marine systems due to natural and anthropogenic stressors. To reach this goal, coastal marine RIs in Europe started JERICO in 2011 as a joint effort to perform a systematic approach to monitor, observe, explore and analyse coastal marine systems in order to reach reliable information of their structure and functioning in the context of global change. This cooperation is being pursued within a cluster of projects:

- JERICO (2011-2015, FP7) focused on harmonising and integrating infrastructure and technologies such as moorings, drifters, ferrybox and gliders and paved the way for JERICO-NEXT
- JERICO-NEXT (2015-2019, H2020) strengthened and enlarged the JERICO network and the interconnection between physics, biogeochemistry and biology.
- JERICO-S3 (2020-2024, H2020) builds on JERICO-NEXT to enhance the current value and relevance of the JERICO-RI, through the implementation of the science and innovation strategy previously defined.
- JERICO-DS (2020-2023, H2020) supports the process towards a structured operational European RI supported by the EU Member States (and associated members) and the EC, and endorsed as a high-value RI at EU level as part of the ESFRI roadmap.

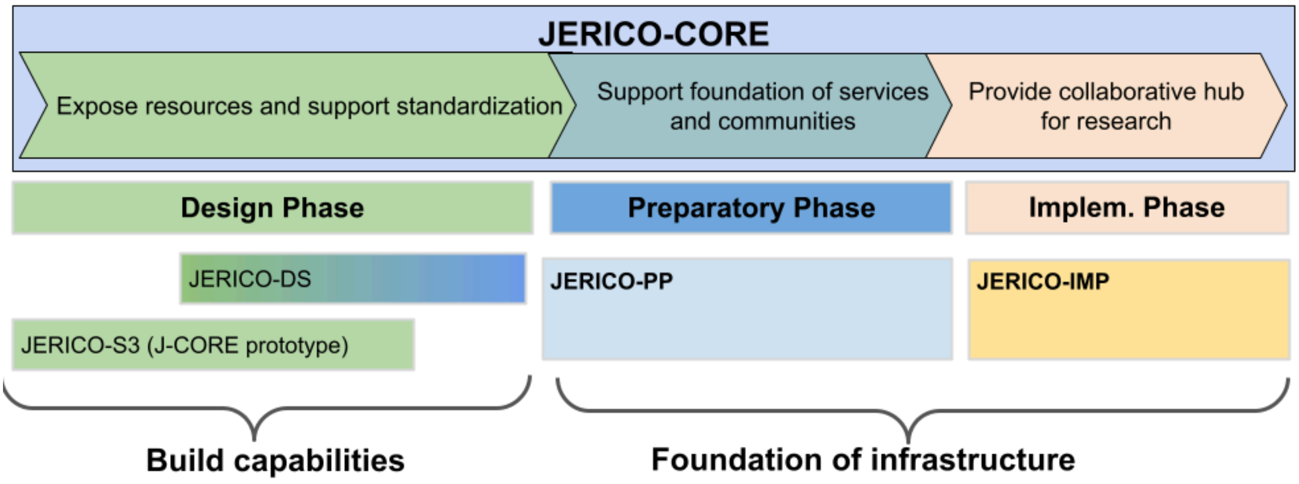
JERICO-NEXT identified the need to create a digital ecosystem that supports the role of virtual access services of facilitating open-access of FAIR and reliable information to research, industry and citizens as well as catalysing international science collaborations. JERICO-S3 is establishing a Virtual Access (VA) framework to allow scalability and discoverability of resources and capabilities of JERICO-RI. It is integrated with advanced services and a VA metrics system that provides an overview of the status and progress of the JERICO digital ecosystem. JERICO-DS is establishing the plans to design and develop a long term sustained Information Management Framework that supports JERICO-RI in the ESFRI roadmap and aligns with the European and global context where cooperation, integration and knowledge interoperability are key to support the work of DTO to empower researchers, citizens, policymakers, and the private alike to visualise and explore ocean knowledge, data, models and forecasts.

Knowledge has become a key aspect to coastal oceanography emerging by the need of accessibility and findability of resources involved in the production of good quality data and models. This is particularly true in the new paradigm presented by the UN Ocean Decade, where DTOs will rely on standard and federated information to virtually represent the ocean for various types of studies and predictions. Therefore, coastal ocean knowledge becomes even more critical in order to better serve the scientific community and address societal and policy needs, finding quickly, easily and automatically the genuine sets of information about coastal ocean assets available among distributed research infrastructures of the JERICO partners. For this reason, JERICO-RI is developing an e-infrastructure that will facilitate the integration of data, data products, facilities and other resources of the coastal ocean and supports the creation of appropriate services for the understanding and protection of the marine environment in European coastal seas. The **JERICO Coastal Ocean Resource Environment** (**JERICO-CORE** or **JCORE**), formerly known as **e-JERICO**, is a unified central hub of JERICO to discover, access, manage and interact with JERICO resources including services, datasets, software, best practices, manuals, publications, organisations, projects, observatories, equipment, data servers, e-libraries, support, training, and similar assets. A knowledge base catalogue of these resources is not a new idea, and has been previously raised in ENVRI-FAIR. JERICO-CORE knowledge catalogue will support services providing JERICO users an optimal way to search and filter valuable resource information to different kinds of users and to gain an integrated form of access to:

- TransNational Access to the JERICO physical infrastructures (platforms and sensors) offered by the national coastal observatories and JERICO technical expertise centres;
- Resources required to both harmonise and implement JERICO data lifecycle management methodologies: Best Practices, tools and services, and e-training modules;
- Quality controlled data that is routinely acquired by the different national coastal observatories, following the FAIR principles;
- Added-value products and services (indicators, nowcasts, analysis, etc.) generated by each individual JERICO thematic expertise centre;
- Dedicated cloud computing resources (hardware and software) allowing researchers to perform advanced analysis on multi-disciplinary, multi-scale, multi-domain and multi-sensor data sets; and
- A community platform which provides the means and the opportunity to share information between all members.

The design of JERICO-CORE is conceived in various phases (see Figure 1) to support the needs of JERICO-RI at its various development stages. The initial design phase encompasses three actions: (1) the creation of a pilot, (2) the analysis of solutions available for a proper design, and (3) the actual design for a long-term sustained e-infrastructure. The creation of the pilot was done under JERICO-S3 to evaluate the feasibility, risk, and performance of JERICO-CORE. It is the foundation to evaluate main applications of a coastal ocean digital ecosystem for successfully supporting the JERICO community as well as to estimate the technologies, resources and time needed for a more sustained solution. JERICO-DS analyses the state-of-art of the coastal ocean and other infrastructures to design a relevant JERICO e-infrastructure that fills the gaps of the current infrastructure landscape and identifies the challenges in creating a solution that supports the goals of JERICO-RI and its User Strategy (*JERICO-S3 - WP9 A sustainable JERICO-RI: Preliminary design towards implementation - Task 9.2: Community of users in JERICO-RI: Analysis of Users & usage strategy*). The preparation phase of the JERICO-RI in the ESFRI roadmap will tackle the proper design for a fully operational JERICO-RI e-infrastructure.

Figure 1: JERICO-CORE Conception Phases



The design phase under JERICO-S3 and JERICO-DS helps building capabilities by exposing resources and supporting standardisation. More specifically, JERICO-S3 defined and developed a VA scalable framework that allows visibility and access of the JERICO-S3 resources and supports the other JERICO VA services with relevant metrics/KPIs to help them evolve. This framework includes the JERICO-CORE pilot and the Virtual Access Metrics System (VAMS). VAMS is a central system to monitor the access to each VA service that is operational under WP11 of JERICO-S3 (D11.1). It was designed and implemented under WP11 to measure performance of VA and assess the fitness for use of their results. These metrics allow researchers to make decisions to fit their dissemination actions to the needs of the data users. The pilot version of the JERICO-CORE was developed under task 7.5 of JERICO-S3 and is being hosted in Datarmor (https://wwz.ifremer.fr/pcdm) and operated by SOCIB (https://www.socib.es/) as two VA services within JERICO-S3 WP11. JERICO-CORE pilot includes a resource catalogue of JERICO-RI resources, a machine-to-machine interfaces to discover and access these resources, a user interface to visualise information, and the integration of four pilot-focused regional thematic services (D2PTS) to demonstrate the benefits of the JERICO-RI information life cycle.

At a more advanced phase of JERICO-RI, JERICO-CORE will be a key platform to support the foundation of services and communities as well as to facilitate collaboration between partners in JERICO-RI, external RIs and other stakeholders. WP3 of JERICO-DS lays the ground to design and build an e-infrastructure that will operate in the long-term, and thus it is scalable and flexible to adapt to new needs of JERICO and the advancements of technologies. The latter requires a comprehensive understanding of the ocean landscape in Europe and globally as well as identifying the strategic collaborations with other infrastructures outside the coastal oceanography. The study of all available solutions and infrastructures will be key to provide the building blocks for a scalable and flexible infrastructure that adapts to the needs of the coastal ocean research community in Europe. Task 3.1 carried out interviews with National Representatives, relevant marine domain Pan-European Infrastructures and initiatives (including cross- and multi-domain) to define what is JERICO-CORE's place and role within the European landscape. WP3 also considers aspects of the infrastructure that complement the technical view of JERICO-CORE (Task 3.3 that includes this milestone report) such as access and security policies (Task 3.2), long-term operations (Task 3.4), management of advanced data products (Task 3.5), monitoring and performance metrics (Task 3.6) and the implementation strategy (Task 3.7). This will result in a more holistic view of an e-infrastructure for long-term support of JERICO-RI.

In this document, we provide the results of the work of Task 3.3. In section 1, we describe the methodology used for the design study of JERICO-CORE. In section 2 and 3, we analyse the European landscape and the requirements collected in JERICO-DS, JERICO-S3 and from various stakeholders are also explained. These led to the elaboration of various conceptual designs at various maturity levels. In section 4, We explain them and the different components of these concepts that will be explored along the design study. In JERICO-DS D3.2 WP3, we investigate the components of the JERICO-DS conceptual design and the available solutions for them. Finally, we provide in section 6 a few possible solutions of technical design based on the combination of technologies previously mentioned. The analysis of these solutions include a summary of the advantages and disadvantages of using them individually and in combination with other components.

3. MAIN REPORT

3.1. *Jerico-DS Technical Design Study in the ESFRI Roadmap.*

The JERICO-CORE Technical Design Study has been prepared keeping special care in following the guidelines and recommendations of the latest **ESFRI Roadmap 2021 Strategic Report** [1] (*Part 2. Landscape Analysis - Section 1.* ***Data, Computing & Digital Research Infrastructures***) , including in the study the recommended e-infrastructures of the major existing European Initiatives (EGI, EOSC, EUDAT, D4Science and Blue-Cloud).

The ESFRI 2021 Roadmap public guide [2], in its *Annex II*, specifies the *Minimal Key Requirements for Scientific Case*, including the *e-needs*, the computer e-infrastructures, communication network systems and any other tools that are essential to achieve excellence in research and innovation.

The next table reproduces the mentioned requirements in each ESFRI phase:

	ESFRI PHASE				
	DESIGN	**PREPARATION**	**IMPLEMENTATION**	**OPERATION**	**TERMINATION**
E-NEEDS	- **Vision on e-infrastructure requirements, including access policy and security measures ready.** - Interfacing with communication networks or distributed calculation or HPC/HTC	- **Conceptual design of e-infrastructure ready.** - Contributions of e-infrastructure resources at all levels (institutional, regional, national, international) described. - **Access policy and Data Management Plan (DMP) outlined.** - Compliance with FAIR principles.	- Technical design of e-infrastructure ready and approved. - Draft operational planning for e-infrastructure service delivery. - Agreements with parties delivering core e-infrastructure services (Central Hub) drafted. - Access policy and DMP approved, including plan for sustainability of data. - Security policy defined and approved. - Implementing FAIR.	- Operational plan ready and approved. - Agreements with service provisioning parties signed. - DMP implemented and security policy deployed. - Operational application of FAIR	- Deployed sustainability data beyond decommissioning.

The present *JERICO-CORE Technical Design Study*, in addition to [*JERICO-DS D3.1 - WP3 - Outlined JERICO virtual resources Access and Security policies*], and [*JERICO DS - WP3 - D3.4 - Data Management Plan for e-JERICO added value products*], establishes the e-needs (highlighted in the table) foundations of JERICO-RI to the ESFRI Roadmap.

3.2. Analysis of the state of the art

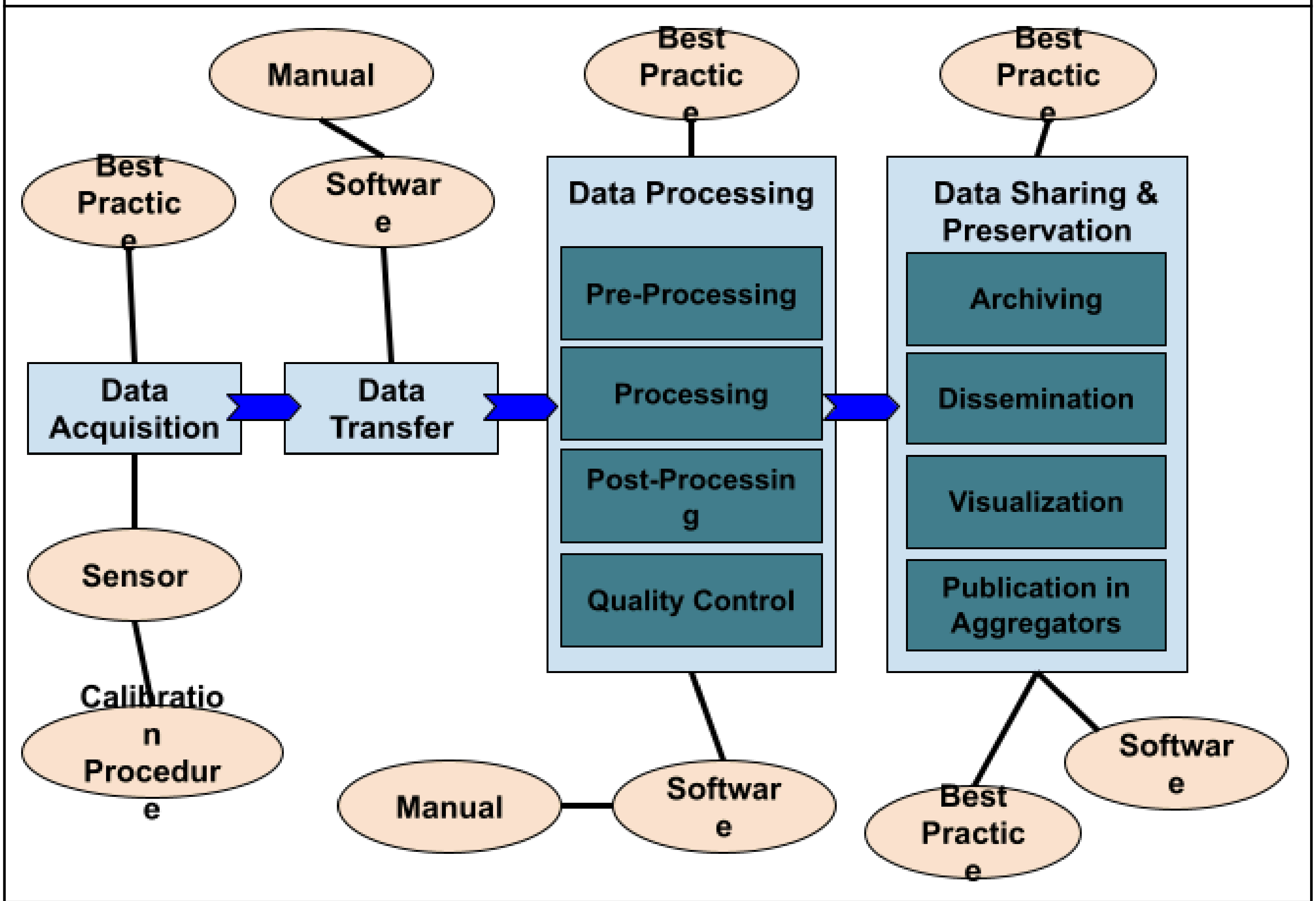
JERICO must align with the current status of data management and programs that facilitate the enhancement and integration of data infrastructures within Europe and at the international level. Therefore, the technical design of JERICO-CORE should consolidate both existing and envisioned technologies, standards, methodologies, and services being offered and used by the European solid digital infrastructures. Blue data infrastructures and initiatives in Europe were created to respond to the initiatives and directives resulting from the guidance of the UN Sustainable Development Goals (UNSDG). They have led the way in Europe to the creation of the adoption of methodologies, frameworks and best practices as well as the development of world class data e-infrastructures of which we can highlight SeaDataNet, EMODnet, Copernicus Marine Service, WEkEO, Blue-Cloud and EOSC. Additionally, the work of various task teams and projects have resulted in the creation of data infrastructures for specific purposes, such (Global) Data Assembly Centers (e.g. Coriolis), databases of specific types of data (e.g. EuroBis) and collaborative platforms for data management and processing (e.g. EcoTaxa). ICT technologies and standards have also benefited from the project and task team work. Their role has been considered in the JERICO Data Management Plan (DMP) (JERICO-S3 D6.1) recognizing the importance of the European data infrastructures and the need to integrate them into the JERICO data flow. JERICO aims to guarantee the collection and dissemination of relevant data according to the 15 principles that aim to make data Findable, Accessible, Interoperable, and Reusable (FAIR). A set of 14 metrics were defined to quantify the level of FAIRness (Wilkinson et al, 2016) and are available at GO-FAIR (https://www.go-fair.org/). They are characterised as follow:

Table 1: FAIR principles

Findable	**F1**. (meta)data are assigned a globally unique and eternally persistent identifier.
	F2. data are described with rich metadata.
	F3. (meta)data are registered or indexed in a searchable resource.
	F4. metadata specifies the data identifier.
Accessible	**A1**. (meta)data are retrievable by their identifier using a standardised communications protocol. – **A1.1.** The protocol is open, free, and universally implementable. – **A1.2.** The protocol allows for an authentication and authorization procedure, where necessary.
	A2. metadata are accessible, even when the data are no longer available.
Interoperable	**I1**. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
	I2. (meta)data use vocabularies that follow FAIR principles.
	I3. (meta)data include qualified references to other (meta)data.
Re-usable	**R1**. meta(data) have a plurality of accurate and relevant attributes. – **R1.1**. (meta)data are released with a clear and accessible data usage license. – **R1.2**. (meta)data are associated with their provenance. – **R1.3**. (meta)data meet domain-relevant community standards.

JERICO-CORE will provide support to increase FAIRness of the data but at the same time needs a minimum degree of FAIR metadata to be functional. JERICO-CORE will link the metadata information of data with the one of the assets that contribute to the creation of data. For example, documents, calibration information, software may all be involved in the data life cycle (see Figure 2). Some of the assets that contribute to the creation of a specific data set in EMODnet will indirectly define the quality of the final data. The way that was followed to calibrate the sensor or the best practice used to deploy the equipments or process the data will provide a hint of the quality that often is necessary for the specific scientific applications. A dataset has likely a higher quality if recognized best practices and procedures were used along the data chain. Therefore, knowing the link between these resources is key to characterise the data that is being analysed. Additionally, an advanced knowledge of the relations between assets allows advanced searching capabilities that may be beneficial for specific purposes. For example, when preparing a Transnational Access (TA) proposal it is interesting to know what observation capabilities a given organisation has to study a specific topic.

Figure 2: Data Lifecycle



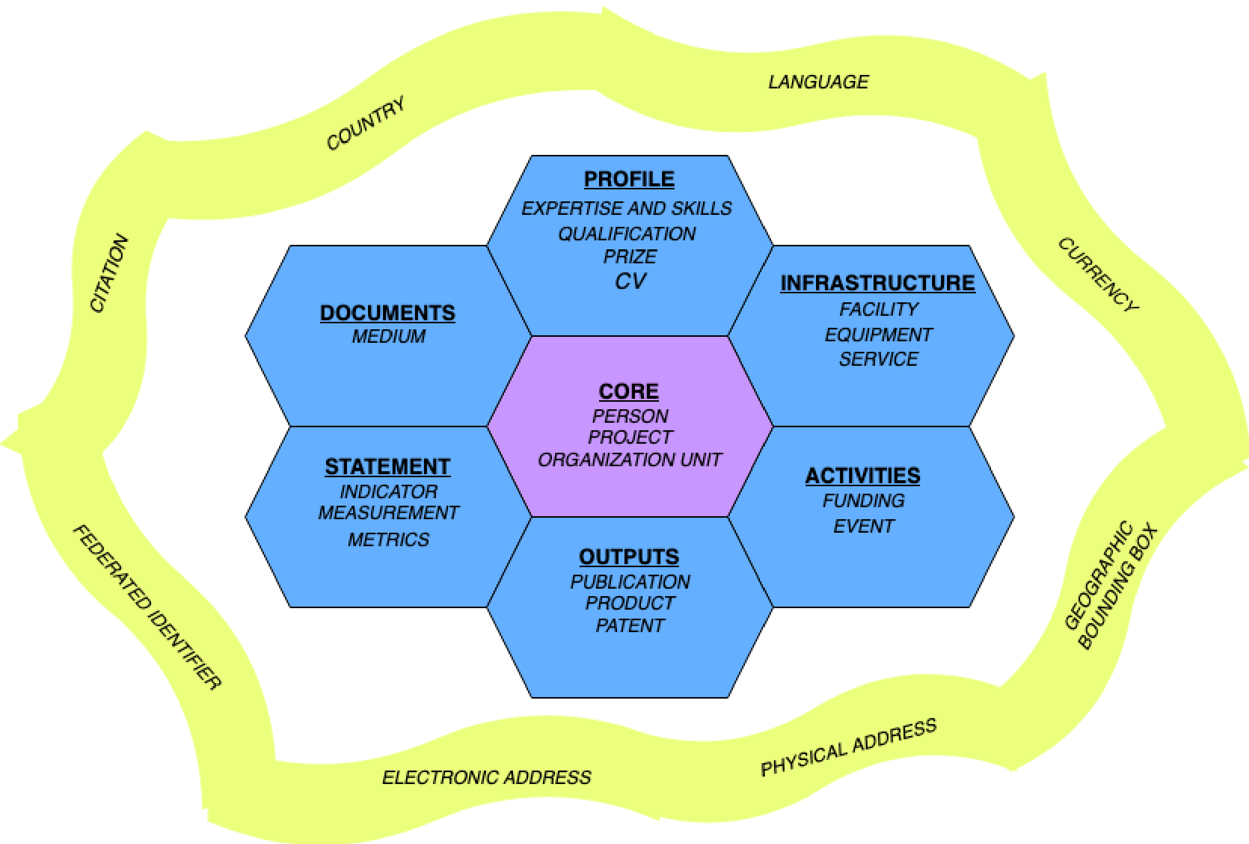
Identifying assets unambiguously allows the linkage between related resources. The use of unique identifiers for data and platforms is well established. The EuroGOOS Data Management, Exchange, and Quality Working Group (DATAMEQ WG) recommends the use of WMO or ICES code to identify platforms. Digital Object Identifiers (DOI) (https://www.doi.org/) should be used for ocean data whenever possible, in particular when used in publications. Unique identifiers should equally be used for other resources for the same reasons. SeaDataNet developed and maintains a series of catalogues that should be widely to provide unique identifiers of resources:

- European Directory of Marine Environmental Data (EDMED): marine metadata formats for data collections

- European Directory of Marine Environmental Research Projects (EDMERP): research projects
- European Directory of Initial Ocean-observing Systems (EDIOS): monitoring programs and networks organisations
- European Directory of Marine Organizations (EDMO): marine institutes

Other identification mechanisms should be used for resources that do not belong to these catalogues. For instance, Open Researcher and Contributor ID (ORCID) could be used to identify researchers or operators. DOIs could be used to uniquely identify software code in a repository or documents. The experience with the pilot study in JERICO-S3 demonstrates that there are many gaps in the universal use of these identifiers and providers use custom ones. In this scenario, names or descriptions are the only available way to link resources implying a very difficult and uncertain result.

Figure 3: Scope of the CERIF data model



The complete set of research should be represented with a data model that formalises the different objects and relationships found in coastal oceanography domain while staying compatible with other research fields, in particular the ones of the ENVRI community. The key to support interpolation is to define a common metadata system that utilises a rich metadata model with formal syntax and declared semantics. Various international initiatives are in place to foster cooperation and knowledge-sharing across the research information community and to promote interoperability of research information. The Research Data Alliance (RDA) was launched in 2013 as a collaborative initiative between Europe, the US and Australia to build the social and technical infrastructure to enable open sharing and re-use of data. There are four core metadata groups of RDA: The groups are: MSDWG (Metadata Standards Directory Working Group), DICIG (Data In Context Interest Group), RDPIG (Research Data

Provenance Interest Group) all coordinated by MIG (Metadata Interest Group). These groups created and endorsed the metadata principles (https://rd-alliance.org/metadata-principles-and-their-use.html). These principles state that metadata should also include information of entities different from data (users, software services, computing resources,…), contextualization (relevance, quality, rights, costs,...) and management (research proposal, funding, project information, research outputs, outcomes, impact…). In line with the RDA principles, euroCRIS (https://eurocris.org) brings together experts on Current Research Information Systems (CRIS) to define data models that interoperate across fields. They defined the Common European Research Information Format (CERIF), a comprehensive information model for the domain of scientific research. Figure 3 illustrates the scope of the CERIF data model by grouping some of the most relevant entities as an example showing the research entities. However, CERIF is one among others of the valid examples of Open Science Graphs information models that are recognized among RDA. Others include Scholix.org, Research Graph, OpenAIRE and PID Graph information models. These metadata schema should be explored from various perspectives:

- Semantic technologies, ontologies, vocabulary (e.g. schema.org, dublin core, DCAT,...)
  - Concept of research entities and their relationships – Specification (Conceptual Level)
  - Description of research entities and their relationships – Model (Logical Level)

- Naming and identification (see Table 2)

- Metadata modelling (e.g. RDF, FRBR, METS) and storage technologies (see ENVRI-FAIR)
  - Formalization of research entities and their relationships – Database Scripts (Physical Level).

- Metadata workflows (when and where to capture metadata during the lifecycle of information).

- Metadata providers and collection techniques (automatic and human oriented metadata) (see Table 2).

- Data and metadata citation.

In relation to providers, we should also consider other sources of information. For example, the OBPS is the reference for best practices and other documents related to oceanography. Github and CRAN are examples of repositories that could provide information on software.

For this study, it is important to consider the existing initiatives around the representation of knowledge in science … ODIS, EOSC, Google, EPOS ENVRI-FAIR knowledge catalogue, EUDAT, OpenAir. Interfaces, ways to represent information and technologies. Ways to interoperate with them.

Table 2: Identifiers and providers at present.

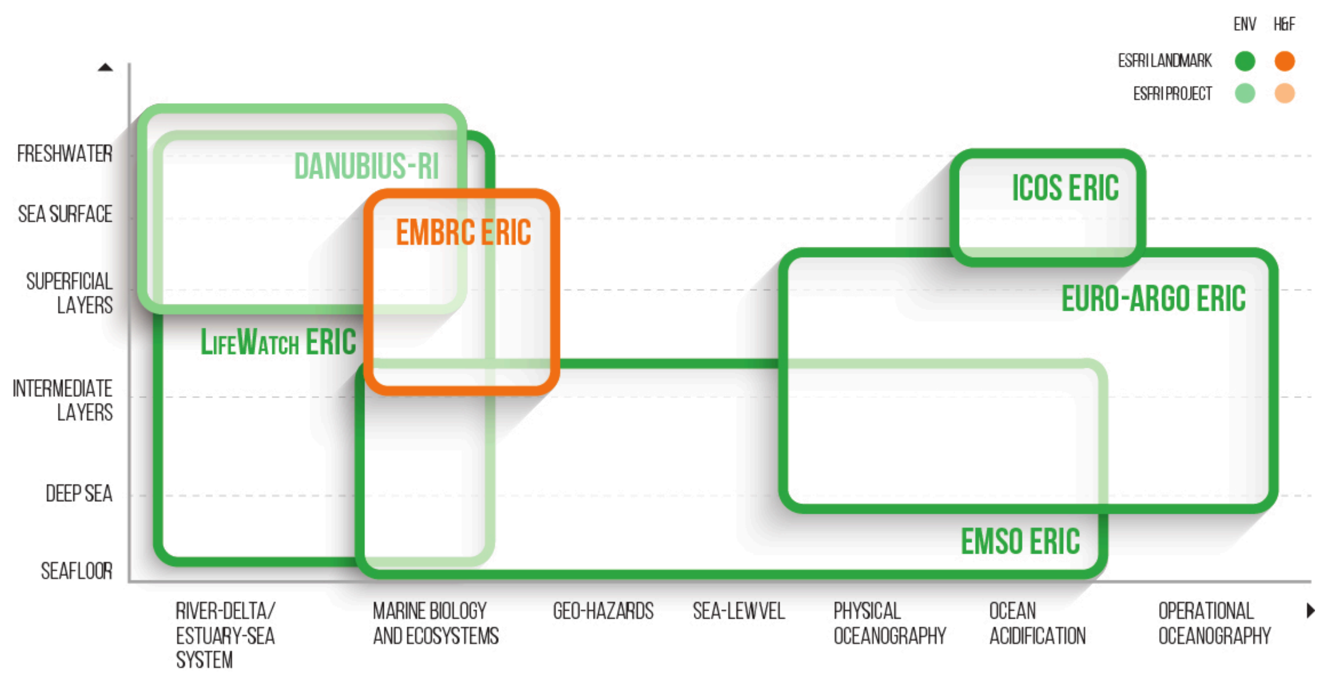
Entity	Identifier	Main Resource Information Providers
Organisation	EDMO	SeaDataNet:EDMO
Project	EDMERP	SeaDataNet:EDMERP
Person	ORCID	Linkedin, ORCID.org,
Publication	DOI	ResearchGate

Best Practice	DOI	OBPS
Manual	DOI	
Calibration document	DOI	
Other Documents	DOI	Aquadocs
Datasets	DOI	EMODnet, SeaDataNet, Copernicus Marine Service, RI servers, SWE sensors
Platform	WMO code, ICES shipcode	Sextant, EMODnet
Deployment	EDIOS	SeaDataNet:EDIOS
Sensor		ESOnet Yellow Pages
Instrument		ESOnet Yellow Pages
Software		Github, CRAN, RI repositories
Service		Manual (json files)?
Events		

3.3. Strategic Collaborations

It is also key to consider what is the role of JERICO in the ENVRI landscape. This will define the interface between JERICO and other key partners and therefore, the technical needs for interoperability with other existing infrastructures. Figure 4 shows the main ENVRI stakeholders that should be taken into account. The ENVRI community [3] created an overview of requirements and standards required for interoperability between the related stakeholders.

Figure 4: Main ENVRI Stakeholders



Focusing on the coastal areas, observations led by JERICO-RI are in-line with the EU Marine Strategy Framework Directive, the EU Water Framework Directive and with regional conventions and Consortiums such as OSPAR, HELCOM, ICES. The added-value of JERICO-RI as a pan European initiative, in supporting the application of these directives and conventions, stands in its capacity to provide high quality integrated data and related services in a harmonised way, as a single entry point. The marine observation landscape in EU [Figure 5a]) is composed of several initiatives and infrastructures, driven by different purposes and with different degrees of integration from national to European dimensions, often with well-established international cooperation. Amongst the marine EU observation and research initiatives, one can list EMSO-ERIC (bottom and water column observatories in open seas), EuroARGO-ERIC, Eurofleet, and the European Ocean Observing System (EOOS), under elaboration, to which JERICO-RI is foreseen to be the coastal component [Figure 4].

Figure 5:

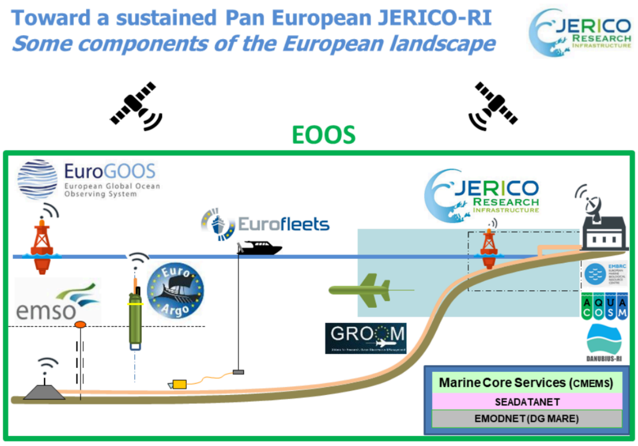


Figure 5a

Figure 5b

In addition AQUACOSM, Danubius, EMBRC ERIC are (or intend to) supplying complementary information and services applicable in the coastal domain [Figure 5b]. Thus, JERICO-RI occupies a specific place in this EU marine landscape, between the open sea RIs and more terrestrial ones. As such, it is targeting to act as the coastal component of the EOOS initiative and to collaborate with surrounding RIs of the Landscape.

3.4. Methodology

This task (task 3.3) considers the key aspects for the design of JERICO-CORE. It is based on the requirements from other tasks of JERICO-DS WP3 outcomes, from the study of significant cases in the precedent JERICO-S3 project [JERICO-S3 D.7.6 - WP7 - "Documentation of JERICO-RI e-infrastructure and capabilities" / 3.4.1. Analysis of existing data and knowledge e-infrastructures / Appendix C: Summary of results of the analysis of the existing e-infrastructures], as well as on the experience with and lessons learned from the design and implementation of the JERICO-S3 pilot. We provide for each step of the study the different aspects to which each of these two projects contributed to. The phases of the analysis is as follows:

- The definition of a Services Architecture that the JERICO-RI community demands. The main requirements (*Appendix A - JERICO-RI Community Requirements*) were gathered, detailed, technically and functionally reorganised and grouped into a set of cooperative interrelated services (*3.5.2. JERICO-CORE Service Architecture*).

- The cloud computing architecture model selected (*3.6.1. The Cloud Computing Model*), also known as cloud computing, offers several significant advantages for organisations and users: scalability, availability and reliability, global access, flexible costs, simplified maintenance, integration and interoperability, etc.

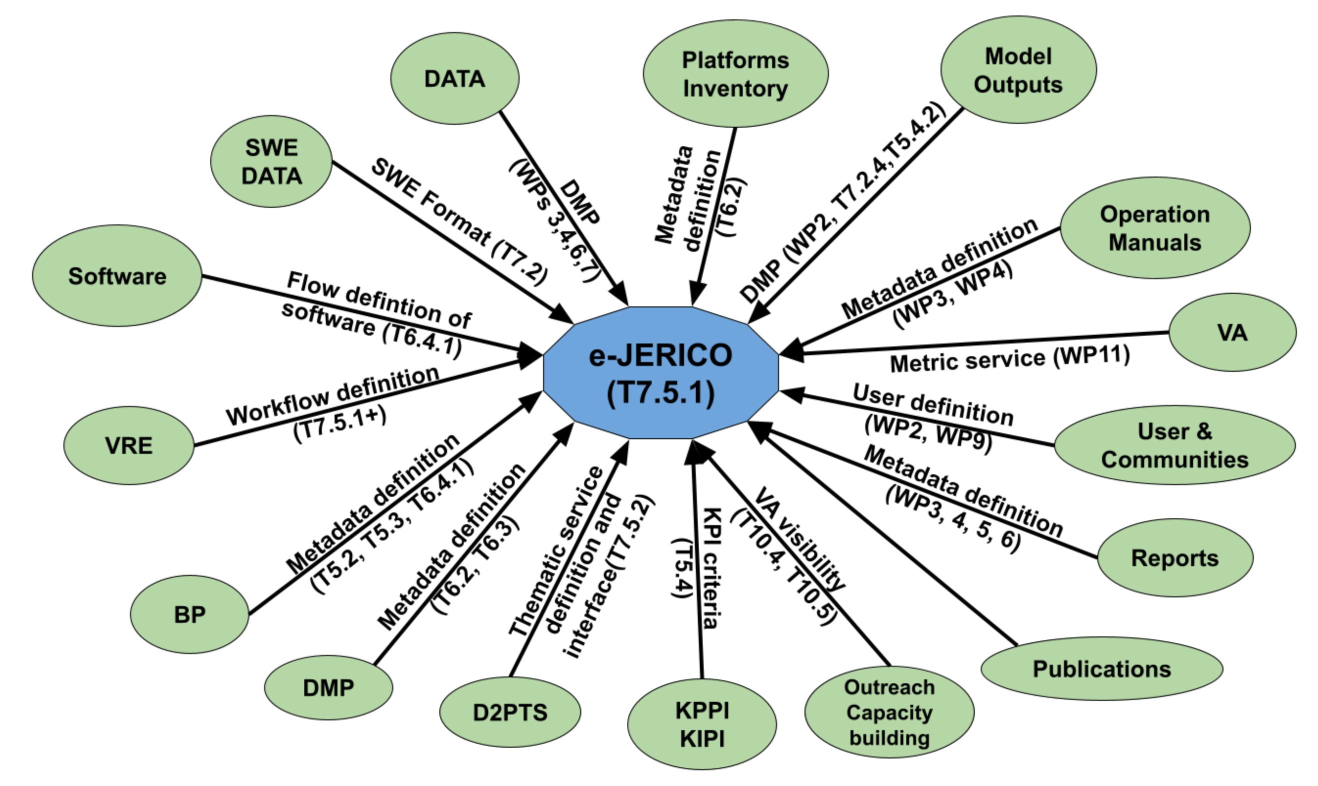
- The study of nowadays existing solutions from the major actual European Initiatives on open science, in order to find the proper combination of e-infrastructures that covers the technical (cloud computing) and functional (architecture of services) for the JERICO-CORE implementation. (*3.7. Implementations solutions*)

3.5. Requirements

Requirements were collected in three different contexts: JERICO-S3, JERICO-DS and conversations with recognized infrastructure and institutions. One of the recurrent requirements was the need to adapt to the existent landscape of infrastructures and methodologies instead of trying to create a similar infrastructure that would have to compete with the already operational and functional data infrastructures. Therefore, integrating to the existing infrastructures (see section 3.2) is one of the main principles that was adopted for the pilot study and that will be considered in this design study.

In the context of JERICO-S3, a set of requirements was collected from the needs of JERICO-S3 project and among JERICO-RI partners, modellers, product developers and other experts in collaboration with all the work packages. Needs from the JERICO-S3 project were extracted from the proposal and associated with functionalities that JERICO-CORE could implement in order to support these requirements. In the context of JERICO-S3 Task 7.5, there were conversations with partners in all work packages. These conversations were done in the context of the pilot but the results are also included here. In particular, it resulted in a conceptual design of the JERICO-CORE pilot that is included in section 3.5. Figure 6 illustrates the interactions with various stakeholders and work package leaders and the outcomes of these conversations. JERICO-S3 D7.6 shows the details of the collection of the requirements and the final results.

Figure 6: JERICO-CORE project artefacts relations



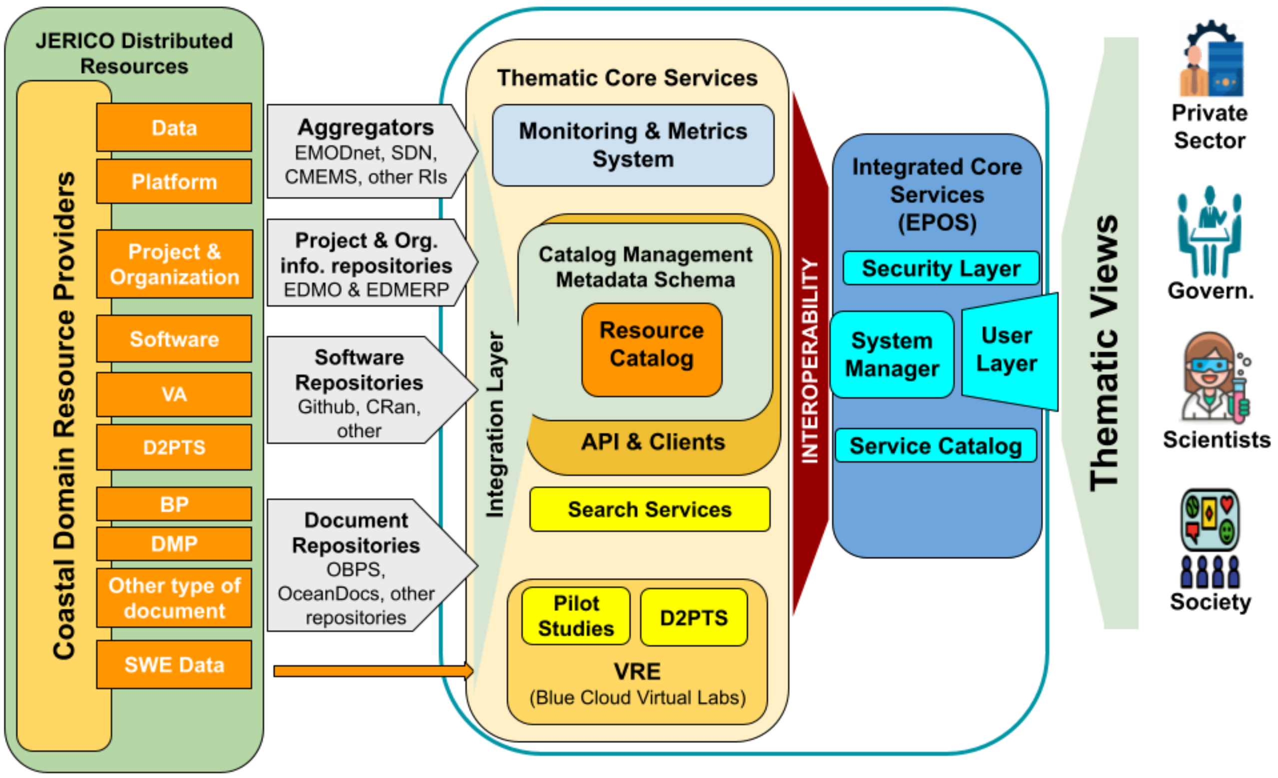
JERICO-DS 3.1 requirements were collected in collaboration with co-leaders of WP1 (scientific) and WP2 (technical). Because of the diversity of needs and ways of work between nations, Task 3.1 carried out interviews and discussions with national representatives, existing relevant marine domain Pan-European Infrastructures (EMODnet, Copernicus Marine Service, SeaDataNet), EU and international initiatives and conventions (EuroGOOS, OSPAR, HELCOM) as well as cross/multi-domain EU infrastructures and initiatives (EMBRC, EOSC, EuroARGO). The results of this work are reported in MS13 of JERICO-DS.

3.6. Concept

3.6.1. Evolution of Concept

There are three conceptual designs with different degrees of maturity according to the phase of the development. At early stages of the pilot, the collections of requirements and the conversations with stakeholders led to the creation of the first conceptual design illustrated in Figure 7

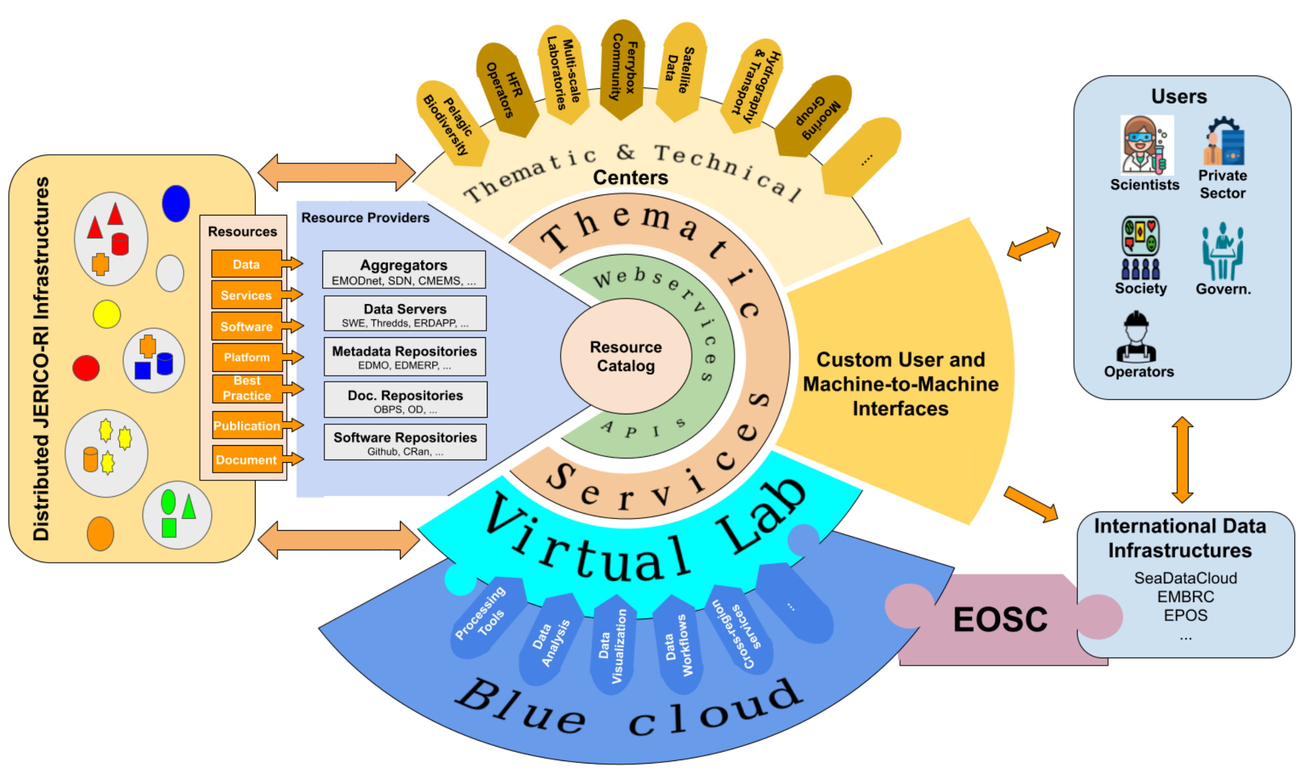
Figure 7: Early JERICO-CORE Conceptual Design



The design was based on the collection of information from the main data providers in order to populate a knowledge base catalogue that will support services developed under JERICO-S3. A series of Memorandum of Understanding (MoU) were agreed to implement this design. A MoU was signed with the European Plate Observing System (EPOS) to use their code to visualise the resources of the catalogue. A MoU was signed with Blue Cloud in order to use a Virtual Lab of their infrastructure to provide VRE capabilities to JERICO-CORE. The concept also includes API services to access the information of the catalogue and VAMS (D11.1) as the metrics system. The integration layer consists of a series of harvesters that use the JERICO-CORE Software Development Toolkit (SDK) [JERICO-S3 D.7.6 - WP7 - "Documentation of JERICO-RI e-infrastructure and Capabilities" - 3.7.1. JERICO-CORE Software Development Kit]. This SDK maps the entities between the representation model of the providers and the model of the resource catalogue.

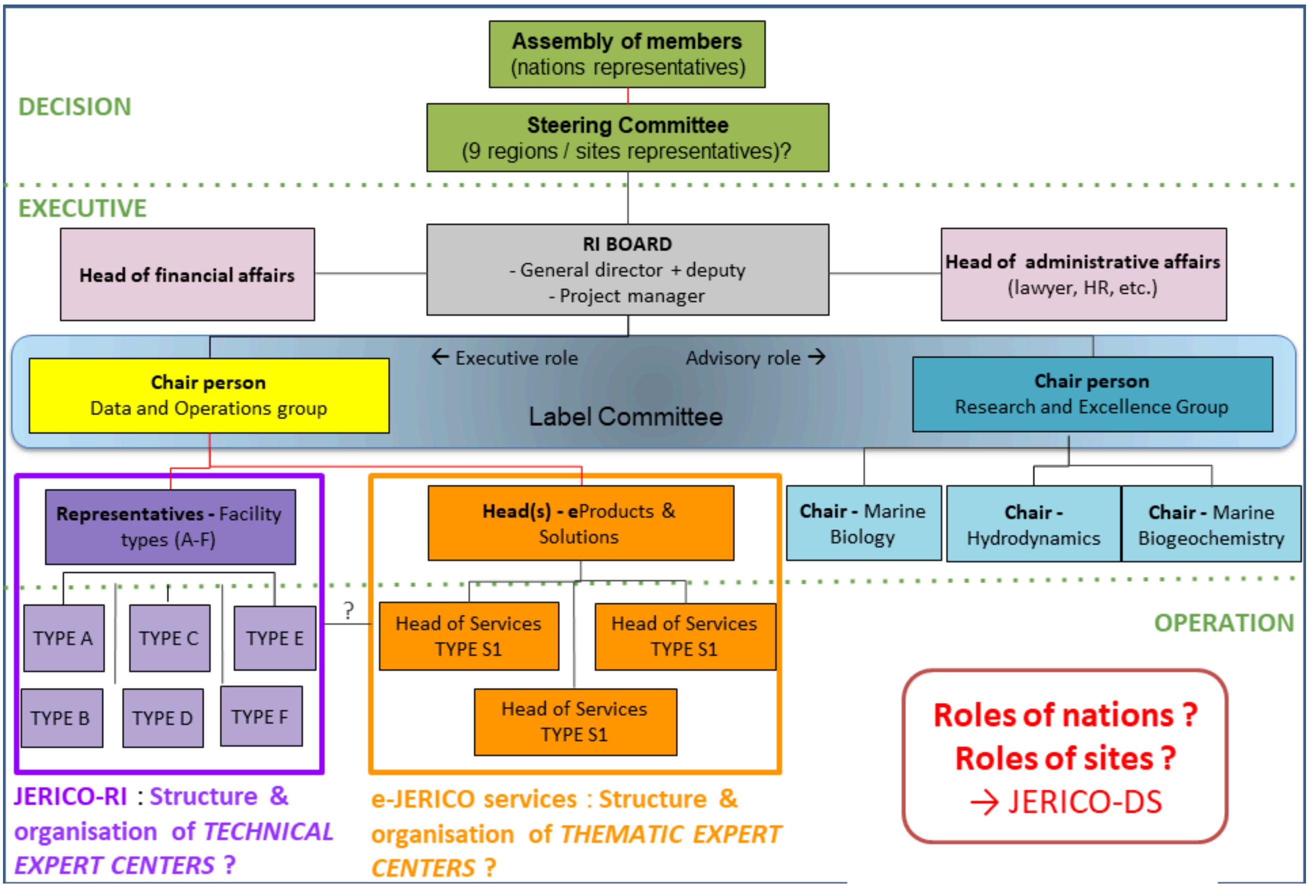
JERICO-CORE aims to provide a coastal ocean resources environment of the JERICO Research Infrastructure that facilitates both the enhanced (virtual) access to all JERICO-RI related resources, and the development of coastal focused services. However, the work of JERICO-DS required at the initial stage a more comprehensive concept that explains the link of the knowledge catalogue with other important JERICO components. This concept was important to start the tasks of WP3 because we required a common initial understanding to facilitate the conversations with stakeholders and when defining requirements, access and security policies, metrics and management of advanced products. We dedicated effort to define the concept of Figure 8 in collaboration with JERICO-S3 and JERICO-DS partners as well as with other external stakeholders.

Figure 8: Advanced JERICO-CORE Conceptual Design



This concept deepens in the connectivity between functional capabilities and services. Underlying the architecture is a flexibility to deal with innovation in science and engineering and meet new demands of users as new capabilities come online. Thus, the system is designed to service current users and support evolution and growth for future needs. In addition to structure, the operating system must have a compelling and intuitive user interface that encourages continuing use. It must also have operational stability and transparency to build and maintain user trust. The architecture and operating environment for JERICO-CORE must deal with the challenge that coastal resources are distributed among JERICO infrastructures using a diversity of standards and implementation methodologies. Resource providers collect these resources following workflows described and accepted in the ocean and coastal domain. However, there is not always a clear connection between the resources that are of a different nature. The JERICO-CORE infrastructure collects information about these resources through the existing providers of data, metadata, web services, software tools, documents and videos, among others. The information of these assets is interconnected in a knowledge base catalogue that is at the core of the infrastructure and offers a broad view on the data as well as observation and operation processes and capabilities in JERICO. The JERICO-CORE inventory represents the relation between the virtual and physical assets from the JERICO Research Infrastructures distributed worldwide among scattered systems. The cross-related nature of the information in the JERICO-CORE inventory facilitates the way resources are found and then accessed at the original source. Therefore, the knowledge of the relation between data, information, metadata, documentation, tools and workflows facilitates the traceability of assets and enhances the FAIRness of the data.

Figure 9: JERICO-RI Project Structure overview



Furthermore, the concept includes a layer of thematic services that will support JERICO by helping the development of customised tools that use the available information. These tools are made available in a collaborative VRE supported by Blue Cloud (https://www.blue-cloud.org). Tools are developed in this framework with direct access to the information and the access of JERICO-RI data, products and services. The access to the resources of the catalogue will be provided via the JERICO-CORE REST API to support machine-to-machine interoperability. JERICO-CORE will support thematic and technical centres that are defined in the JERICO-RI structure (see Figure 9). In the context of the EOSC, collaborative developments and improved access to resources, tools and services in a common infrastructure will enhance the collaboration with other research infrastructures. This two-way interoperability will benefit both JERICO and its stakeholders.

These resources and services will be accessible to scientists, private sector, society, government and technical operators through customised user interfaces to assist their needs. Additionally, external users and international data infrastructures will benefit from machine-to-machine interfaces to automate the access to JERICO assets. Machine-to-machine access to services and underlying resources will enhance the interoperability of JERICO-RI with other international infrastructures and allow research across domains and fields.

The motivation for JERICO-RI is that many coastal ocean problems are complex, requiring experts in differing disciplines to co-design and implement solutions. JERICO-CORE aims to support JERICO-RI in this regard. At the simplest scale, observers and data managers will work together to deliver data and information in consistent and easily readable formats. At a higher level, monitoring and understanding of the coastal physical oceanography, where, for example, examining the water mass flows around the Iberian Peninsula, requires multiple sensors, observations and a range of expertise. As we approach the highest level, such as sustainability of marine biodiversity, oceanography expertise and data in physics, biogeochemistry, biology and ecosystems is needed over multiple times and scales. The services described in the next section illustrate a range of capabilities offered by JERICO-CORE

The third conceptual model was defined in the context of Task 3.3 of JERICO-DS. It expands the previous conceptual framework of the JERICO resources by including other concepts that respond to the needs expressed by the community. The JERICO community, the Organizations that comprise it, consists of scientists and technicians focused on the observation and research of the repercussions of the rapids and continuous changes that suffer the natural resources of our sea coasts.

Our community of scientists, through the requirements collected in the JERICO DS's MS13 - "e-JERICO requirements compiled", expresses its concerns about the need for technical means to deal with these changes in a quick and efficient manner. The JERICO Community requires a reliable knowledge-based platform to unify data, analyse, process, and communicate the results of coastal resource observations from various technical infrastructures. This platform will enable quick and efficient studies and decision-making based on the observations of our researchers. By providing timely and high-quality results, our coasts' resources will benefit from this platform.

In addition to the excellent scientific-technical infrastructures and researchers of JERICO-RI in the field, the e-JERICO e-infrastructure stands out with a unique and distinctive approach compared to other platforms. It focuses on developing a robust resource catalogue managed through versatile harvesting processes. This allows users to easily locate and access a wide range of federated data, information, and certified high-quality resources.

3.6.2. JERICO-CORE Service Architecture

The consequence of the conceptual evolution of the JERICO-CORE design, as described in point *3.5.1. Evolution of Concept*, is the gradual change of the approach of the JERICO technical design. Along these progressive variations, JERICO-CORE concept was always based on a distributed architecture of interconnected components which is supported by a reliable, flexible and powerful source of information stored in a central resource catalogue.

Once this degree of maturity has been reached, JERICO-DS aims to study the existing technological solutions and architecture that can be used for the design of the desired JERICO-CORE system. This will take into account all the main requirements of the project, collected in MS13 of JERICO-DS, and JERICO-S3's D7.6, resulting in a new evolved conception of service architecture.

A requirements/service component matrix is presented in Appendix A, in which each gathered requirement is related to the service component(s).

Figures 10, 11 show the new *Service Architecture* from different points of view. Figure 10 places the new JERICO-CORE architecture as a central element that communicates the rest of components of the JERICO-RI infrastructure. This vision brings the JERICO Community together around its new e-infrastructure.

On the other hand, Figure 11, presents the technical view (albeit very simplified) of the elements of the service architecture, with the main elements of the application platform intended to support the upper layers of services.

In the following subsections, the function and relevance of each service component will be explained.

Figure 10: JCORE - Service Architecture Landscape View

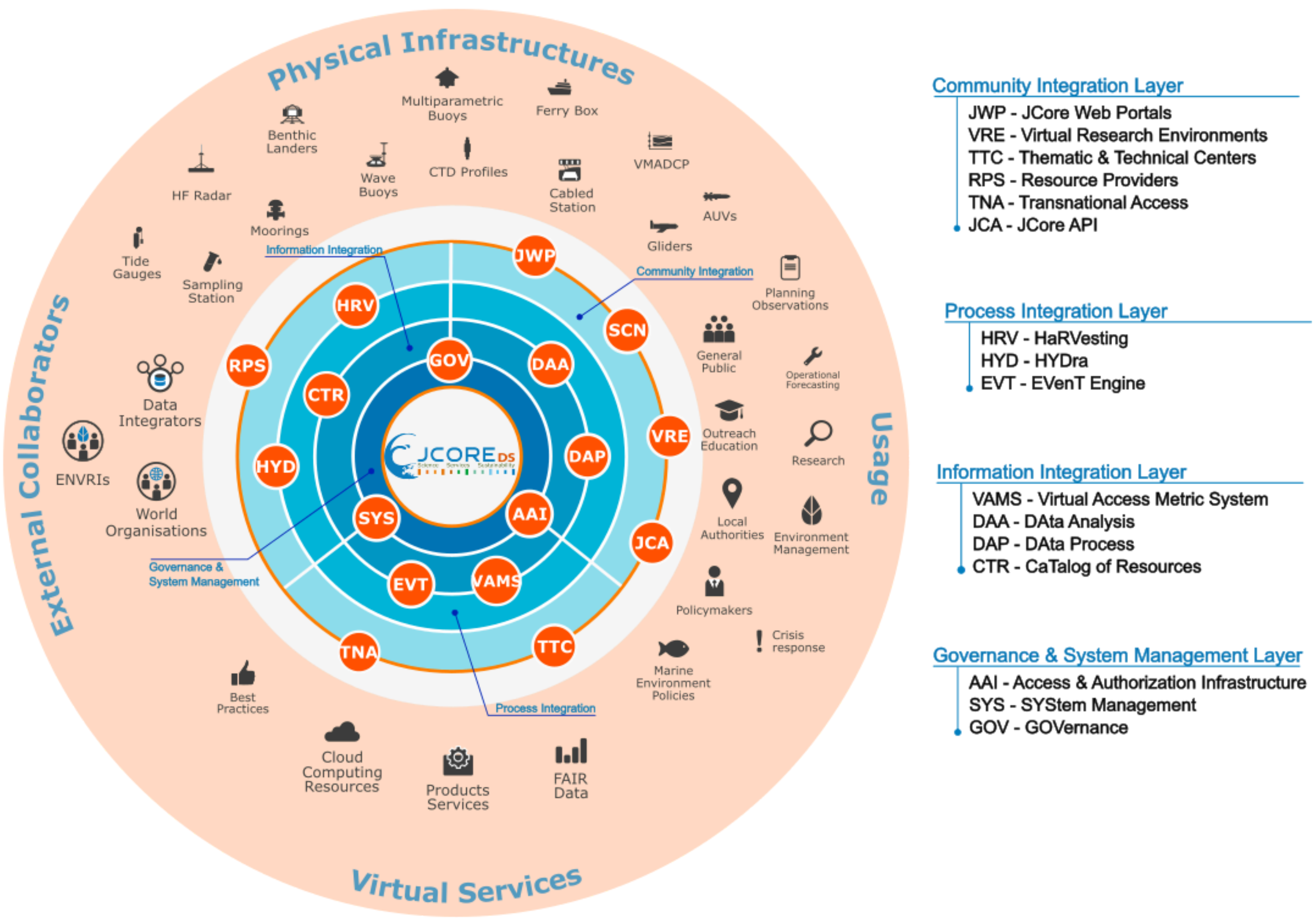
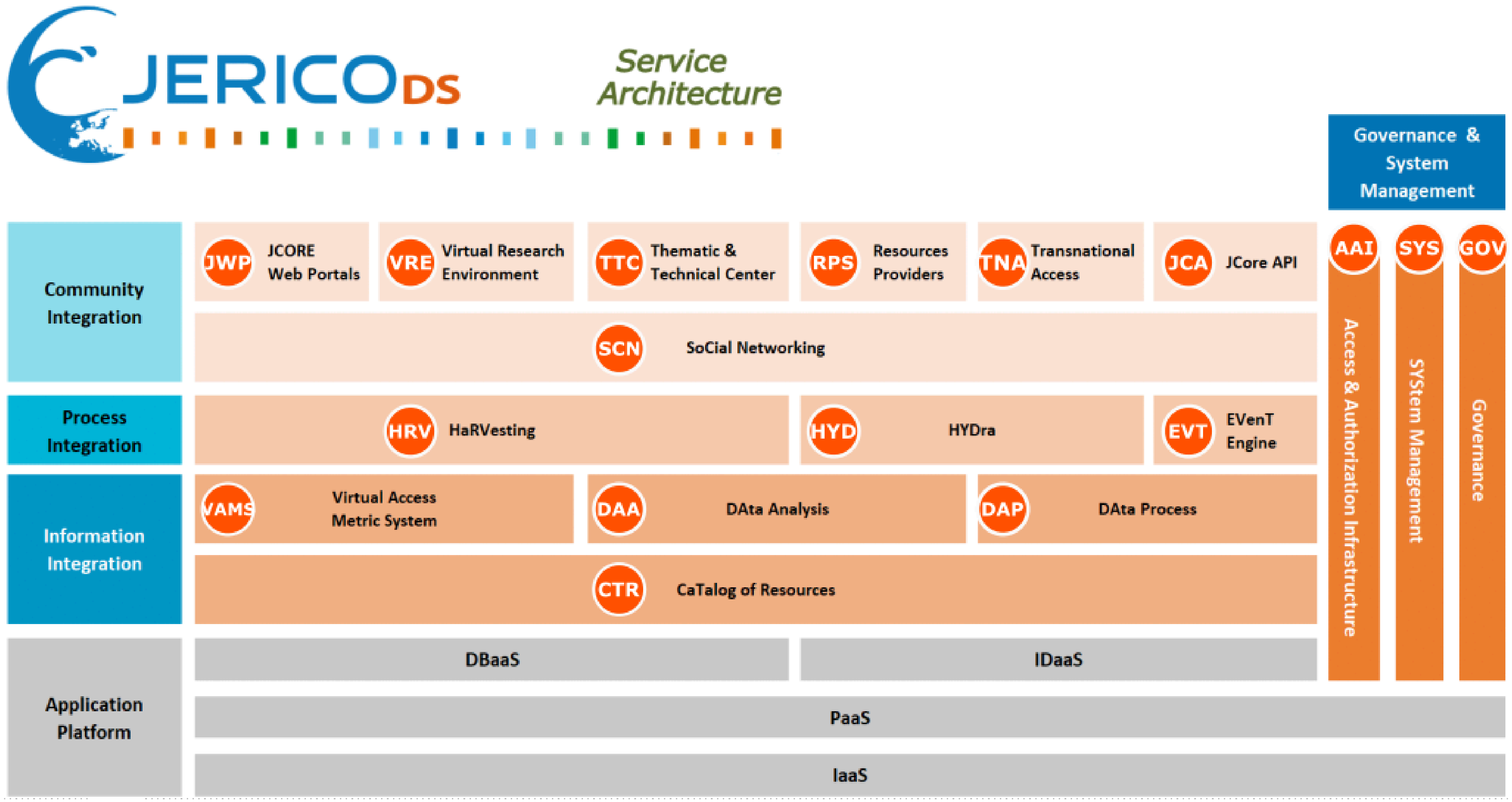


Figure 11: JCORE Service Architecture - Stacked View



3.6.2.1. Community Integration Layer

One of the main targets of JERICO-RI infrastructure is the openness to additional communities. The Community Integration Layer is the key component in both aiding new communities to prepare themselves to participate in the JERICO-RI infrastructure, and growing the adoption and use of JERICO-RI services.

This set of components promotes standardisation of data access and exchange tools and makes it easier to contribute to JERICO-RI services, tools and interfaces. These activities could provide Best Practices that may be adopted by other communities and thus, simplify the operations by reducing the number or types of services requested by different communities.

The Community Integration Layer will address access and security policies related to legal and ethical issues of all nations partners regulations, in order to be compliant with laws in confidentiality and data protection.

The Community Integration Layer establishes operational, service, and end-user agreements that clearly define roles and responsibilities in use, monitoring, management and governance, and also the scope of services and levels of performance.

The components of this layer are:

JCORE - Web Portal Services. The primary role of the Web Portal Services is to be the website that serves as a centralised gateway or entry point designed to provide users with a unified and convenient interface. Users will be able to search for specific marine data, explore interactive visualisations, access advanced analytical tools, and collaborate with other researchers and stakeholders in the coastal marine domain. The JCORE Web Portal Services wil promote interdisciplinary research and knowledge exchange by connecting different communities and fostering collaborations.

The JCORE - Web Portal Services plays a crucial role in providing researchers, policymakers, and other stakeholders with a centralised platform to access and leverage marine data, services, and tools. It supports the advancement of marine science and contributes to a better understanding and management of coastal marine ecosystems and resources.

VRE - Virtual Research Environment Services. A VRE is not only a subsystem in JERICO-CORE. VRE is the short term for the set of tools and technologies used by researchers to perform their activity, interact with other researchers and to make use of distributed computing resources running over technical infrastructures connected by world wide networks to facilitate and enhance the processes involved in research.

The concept of VRE is not generic, fixed and immovable. The VRE also include the context in which VRE's tools and computing resources are used. The JERICO VRE Services are even more than a generic VRE: They will be unique in the sense of they will be endowed with the discipline-specific tools (such as document hosting, data analysis, visualisation, simulation, best practices, teaching tools,etc.), for the research based on the coastal observing systems.

TTC - Thematic & Technical Center Services. JCORE platform, through the TTC Services, offers all interconnection functionalities for the publication and management of information of the different groups of multidisciplinary facilities of European coastal resources research infrastructures. These services are one of the main direct data media between the Community of JERICO researchers and the rest of the JCORE system. The expectations put in these services are very high and this project not only marks the goal of obtaining the massive and great quality of the data obtained in the field of investigation, but also to achieve it in real time.

TTC Services also aims to be the tool that coordinates research infrastructures in international size projects from which data and models will be obtained, and will demonstrate how climate change problems can be observed in a coordinated and effective way, saving efforts and resources between nations, and giving a global perspective in short, medium and long time.

RPS - Resources Providers Services. These services are intended for information suppliers external to JERICO-RI such as the main information aggregators currently existing, and who wish to provide data and actively participate in the JCORE platform. RPS are conceived as a set of interfaces and tools that allow to Resource Providers to become information provider, but will also benefit in the same way that the Thematic & Technical Centers of JERICO would do, receiving feedback on the use and importance of their information in the JERICO community, gaining visibility on an exclusive platform, and moreover, be rewarded with a metadata assessment service.

TA - TransNational Access Services. The aim of the JCORE's TNA (or TA, TranNational Access) Services tool, is to make JERICO's research platforms and facilities visible to the JERICO Community through the Internet, so that they can be reserved and used in order to cover the maximum possible number of research projects with the minimum possible actions, which will allow a greater and better reuse of these resources.

Through this tool, all groups of researchers will announce to the rest the calendar of projects and interventions and therefore will offer the opportunity to reserve and combine resources with other researchers through an agile and intuitive communication tool, which will know and include perfectly the life cycles of the research outcomes (e.g. datasets) of different coastal resources, and with which the community of JERICO will reduce research investigation costs.

JCA - JCORE API. The majority of the public APIs only offer outgoing information to external third -party systems with which they connect. JCORE API, in addition to a valuable source of information for other platforms, will be a novel interaction tool between systems that, thanks to the use of open standards for the exchange of information, will allow third parties to create their own spaces within the JCORE platform, taking advantage of all knowledge and experience of the JERICO community in the form of a wide catalogue of resources, valuable FAIR data, semantic technologies and even real-time access to JERICO Research Infrastructures.

SCN - SoCial Networking Services. SCN is the government tool of the different services that form the Community Layer: It is the manager of the social environment of JERICO, capable of creating and managing communities of users, offering services of communication, meeting and dissemination (newsletters, broadcasting, events, multimedia streaming, online meetings, discussion forums, etc.), which will also be reflected in the rest of JCORE Web Portal Services, giving visibility to JERICO Community actions.

These services represent for the Community of JERICO a valuable tool for coordination and internal communication for its members but also represent a unique opportunity (non-existent as such today) to share their work with the rest of researchers, educators, policy makers, local authorities, Governments, and the rest of the general public, promoting the dissemination and investigation of marine coastal resources with the objective of raising awareness among them, about their precariousness, but at the same time offering a valuable instrument of communication to support decision making.

3.6.2.2. Process Integration, Automation, and Optimization Layer

This layer is a framework for process integration of functional applications, automation for operations workflows, and additional functionalities for processes optimization. The proposed approach combines applications and workflows using integrated process/data models and forms a foundation for processes optimization.

The components of this layer are:

HRV - HaRVesting Services. Thanks to the large amount and variety of information on the Internet, it is today the quintessential resource where researchers from all over the world can exchange the results of their research activity. And although it is easy to find information on any topic on the Internet, what is really difficult is to automate the process of extracting information accessible online, due to the great variety of data structures and the great diversity of information sharing methods. There are many different proprietary formats, standards and protocols that make the task of data harvesting and data mining really difficult.

This is where HaRVesting Services become a fundamental piece in the service architecture. Obtaining, curating, refining, and classifying information is a challenging and demanding process that requires a complex and well coordinated set of scalable software technologies, ranging from automated and well orchestrated and monitored workflows, to complex machine-learning algorithms able to understand the volatile information life-cycle.

The HaRVesting Services will generate interrelated and indexed contexts based on ontologies, feeding the Resources Catalogue with all the possible wealth of interconnected information of the JERICO resources.

HYD - HYDra. Sensor technology, computer technology and network technology are advancing together while demand grows for ways to connect information systems with the real world in real time. In this sense, given the heterogeneous diversity of technology that conforms the physical monitoring landscape of any research infrastructure, it is clear that JERICO-RI is a great integrated pan-European multidisciplinary and multi-platform research infrastructure dedicated to a holistic appraisal of coastal marine system changes.

HYDra Services are the ambitious piece of middleware that will connect JERICO-RI with the JERICO-CORE platform, opening a wide range of possibilities for the JERICO Community due the connection with a wide infrastructure of sensors, actuators, and monitoring devices in general, which are placed in a vast pan-European landscape.

HYDra Services are now conceived as the Digital Model that connects the Physical System and the Digital System of JERICO; but in the future, in a first short-term phase, HYDra Services will be evolved to gain real-time (i.e. SWE) connection between Physical and Digital worlds. This evolution will transform JERICO from the actual *Digital Model* to a *Digital Shadow*. But finally, in a mid/long-term, HYDra will be able to establish a bi-directional connection, giving the possibility to the Digital System to act on the Physical System, evolving the JERICO from a *Digital Shadow* to a *Digital Twin CPI* (Cyber Physical Infrastructure).

The aim of a Digital Twin CPI is to have a more intelligent and reactive environmental monitoring system, and it is able to decide, and also order how the physical infrastructure should act to confirm and prevent environmental risks, adding value and increasing efficiency and effectiveness.

EVT - EVenT Engine. In a vast and decoupled but cohesive cloud infrastructure, an asynchronous, resilient, and flexible means of communication is necessary to allow all the components to be asynchronously communicated and coordinated.

The EVenT Engine Services will support communication between JERICO-CORE components in such a way that not only machine-to-machine communication will be possible, but also the possibility of communication and interaction between the members of the JERICO Community will be opened. It would be possible, i.e., for the administrator users of the SoCial Networking Services to send broadcast messages about events and news to the Jerico Web Portal components in a simple and unified way. Researchers will be able to receive and send messages, news, publications, etc. easily and transparently. The HYDra services will be able to communicate to the DAta Analysis and DAta Process components the arrival of information from sensors. These are simple examples of the need for these event services, but if we also take into account the possibility of opening them to the outside through APIs, then we have the perfect platform feedback system with outside incoming information.

3.6.2.3. Information Integration Layer

VAMS - Virtual Access Metric System. Access metrics to Virtual Access (VA) services are an essential piece of information to manage a research infrastructure (both at the strategic and planning levels). They allow managers to make informed decisions and prioritise actions in a data-driven way. In a world of open science, where the performance of researchers or research institutions cannot be measured by bibliometric data only, access metrics to a VA service are important additional indicators. [4]. As a matter of fact, the OECD has developed a set of metrics to [5] assess the overall impact of research infrastructures. Some of them are related to data access: S6-Number of scientific users, T27-Data sharing (Number of data requests, Number of data accesses).

Access metrics are quantitative assessments used for assessing user access to research resources and provide a first insight into their level of dissemination. The role of access metrics is not only to measure the research performance as they also allow researchers to assess the fitness for use of their results. Thus, closing a feedback loop that allows researchers to make decisions to improve their dissemination to the needs of the data users. [6]

DAA - DAta Analysis & DAP DAta Process. Although split as two kinds of services, it has no sense to explain each one separately, due the tightness of their functionalities. DAA & DAP will be made up of sets of open source algorithms with well-defined interfaces, in such a way that they can be sequenced within execution pipelines, providing a wide variety of combinations of data processes.

DAA & DAP Services will be the basis for the harmonisation, transformation and generation of data products that cover a large number of requirements: from the processing and analysis of information coming from sensors, to the generation of data products that, well orchestrated and connected to the EVent Engine Services, can be used to create reports, dashboards, automatically publish information in viewers embedded within web pages, give relevant information to create alarm systems for the prevention of adverse climatic phenomena, etc.

The DAA and DAP algorithms will be in private spaces within the system (preserving the security and integrity of JERICO-CORE's internal processes), but they will also be shared through replication in public spaces accessible from the VRE services so that the entire JERICO Community can use them in their own research processes.

CTR - CaTalog of Resources. Although the Catalog Of Resources in the pilot of JERICO-S3 was a Resource Description Framework (RDF) database, it is clear that an index of resources is not enough to support all the functionality required by JERICO-CORE: there exists the need to collect and process data products.

The management, processing and, in many cases, storing of such datasets is beyond the capacity of traditional technological approaches based on local, specialised data facilities. Such data are characterised by the well known three V's of Big Data paradigm: Volume, the data dimension; Velocity of data collection, processing and consumption that is demanding in terms of speed; and Variety, the data heterogeneity in terms of data types and data sources requiring integration, is high.

For these reasons, the JERICO-CORE will incorporate a new Catalog Of Resources built on a Hybrid Data Infrastructure (HDI).

A Hybrid Data Infrastructure is a new type of Data Infrastructure specifically conceived to deal with data-intensive science (e-Science), but incorporates a powerful RDF component, such as JERICO-S3 Catalog of Resources.

3.6.2.4. Application Governance & System Management

This layer contains the cross-cutting services that maintain and manage the rest of the services, not only from a purely ITIL (Information Technology Infrastructure Library) point of view, but also from a quality management point of view of the scientific production collected in the platform.

This layer of services is what highlights the value of human, technical and scientific capital, specialised in the management of cloud platforms and in the research of coastal resources.

AAI -Access & Authorization Infrastructure. A security topology describes an abstract layer of AAI's components, requirements, and trust relationship between them. It preserves the provisioning of AAI across different environments for interoperability, portability, and enables a dynamic trust relationship with other services on demand.

Conceived as a Federated Identity Management Infrastructure [6] based on the IDaaS security model, the *AAI Services* decouples the authentication and authorization tasks from the rest of services as a manageable resource, composed by a federated cloud of Identity Providers.

Since Identify-as-a-service (IDaaS) model controls the complete security chain, it can coordinate automated trust negotiation between cloud services in federated security domains, preserving user's privacy.

GOV - Governance. In JERICO-DS, the Governance Services offer to authored members of the JERICO Community to create the best fitted and automated policies and principles that establish the appropriate degree of control around the data-lifecycle, in order to offer the better FAIR hallmark.

SYS - SYStem Management. Cloud computing management and governance is broadly known as a view of IT governance, focused on accountability, defining decision rights and balancing benefit or value, risk, and resources in an environment embracing cloud computing [7] [8] [9].

The SYStem Management will be a cloud computing system-monitoring and governance framework that creates business-driven policies and principles that establish the appropriate degree of investments and control around the lifecycle process for cloud computing services .

3.7. JERICO-CORE Technical Design Study

Whenever a community starts developing an IT solution for its use case(s) it has to face the issue of carefully selecting "the platform" to use. Such a platform should match the technical and functional requirements and the overall settings resulting from the specific application context (including legacy technologies and solutions to be integrated and reused, costs of adoption and operation, ease in acquiring skills and competencies, etc.). There is no one-size-fits-all solution that is suitable for all application contexts, and this is particularly true for scientific communities and their cases because of the wide heterogeneity characterising them. However, there is a large consensus that solution approaches based on ad-hoc and "from scratch" development of the envisaged supporting environments are neither viable (e.g. high "time to market") nor sustainable (e.g. technological obsolescence risk).

Cloud computing solutions have been increasingly adopted by users and providers to promote a flexible, scalable and tailored access to computing resources. Cloud computing is now being endorsed by organisations with limited resources or with a more articulated, less direct control over these resources. The challenge for these organisations is to leverage the benefits of cloud computing while dealing with limited and often widely distributed computing resources. This study focuses on the adoption of cloud computing for JERICO-CORE. The proposed solutions leverage a federated approach to cloud resources in which the JERICO Community access JERICO-CORE services through multiple and largely independent cloud infrastructures. This approach allows for a uniform authentication and authorization infrastructure, a fine-grained policy specification and the aggregation of accounting and monitoring. Within a loosely coupled federation of cloud infrastructures, the JERICO-RI users can access vast amounts of data without copying them across cloud infrastructures and can scale their resource provisions when the local cloud resources become insufficient.

This section describes the importance of the cloud computing model, and the nowadays main federated infrastructures based on this model.

3.6.1. The Cloud Computing Model

Cloud computing is a model for enabling a convenient, on-demand network access to a massive shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.

3.6.1.1 Cloud Computing - Essential Characteristics

- ***On-demand & self-service***. A user can provision, automatically and unilaterally, computing capabilities, such as server time and network storage as needed, without requiring direct human interaction with each service provider that is part of the cloud computing platform, without needing to worry about the underlying infrastructure.

- ***Broad network access***. Resource capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous client platforms (e.g., workstations, servers, mobile phones, tablets, laptops, IoT devices, etc.).

- ***Resource pooling***. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. Examples of resources include storage, processing, memory, and network bandwidth. With the resource pooling technique, there is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter).

- ***Rapid elasticity***. Resources can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

- ***Measured service***. Cloud systems automatically control and optimise resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilised service.

3.6.1.2 Cloud Computing - Service Models

- ***Software as a Service (SaaS)***. The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure . The applications are accessible from various client devices through either a thin client interface, such as a web browser (e.g., web-based email), or a program interface. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user specific application configuration settings.

- ***Platform as a Service (PaaS)***. The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly configuration settings for the application-hosting environment.

- ***Infrastructure as a Service (IaaS)***. The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components (e.g., host firewalls).

3.6.1.3 Cloud Computing - Deployment Models

- ***Private cloud***. The cloud infrastructure is provisioned for exclusive use by a single organisation comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organisation, a third party, or some combination of them, and it may exist on or off premises.

- ***Community cloud***. The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organisations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organisations in the community, a third party, or some combination of them, and it may exist on or off premises.

- ***Public cloud***. The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organisation, or some combination of them. It exists on the premises of the cloud provider.

- ***Hybrid cloud***. The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardised or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

In its early stages, "the cloud" was a term used to express the empty space between the end user and the provider. In 1997, Professor Ramnath Chellapa of Emory University, defined cloud computing as the new "computing paradigm, where the boundaries of computing will be determined by economic rationale, rather than technical limits alone." This somewhat ponderous description rings true in describing the cloud's evolution. The cloud gained popularity as companies and organisations gained a better understanding of its services and usefulness for the Science and Research Community. In the next section, the main cloud computing platforms created under the umbrella of the European Commission initiatives to promote scientific excellence, will be analysed as possible solutions to implement the JERICO-CORE solution.

3.8. Implementation Solutions

According to the cloud computing model selected to implement JERICO-CORE, in order to not reinvent the wheel, and taking advantage of the major actual European Initiatives on open science, this section analyses and compares the main e-infrastructures where JERICO-CORE can be fully or partially implemented.

3.7.1. EGI

3.7.1.1. Introduction

EGI (European Grid Infrastructure) is a publicly-funded e-infrastructure based on a federation of national e-infrastructures (in Europe and beyond), known as EGI resource centres, that provides advanced computing and data services for research and industry. [9]

The EGI resource centres rely on the expertise of the 'EGI Foundation', a not-for-profit organisation that coordinates EGI. The Foundation oversees areas such as infrastructure operations, user community support, contact with technology providers, strategy and policy development, flagship events and dissemination of news and achievements. The EGI collaboration has participants and associated participants drawn from representatives of national compute/data centres (so called NGIs – Next Generation Infrastructures) and Intergovernmental Research Organizations, such as CERN. These entities operate the physical resources and services that comprise the 'EGI infrastructure' and the 'EGI service catalogue'.

The EGI Cloud service is implemented in the form of a Federated Cloud. This EGI Federated Cloud is a standards-based, open cloud system federating institutional clouds in order to offer a scalable computing platform for data and/or compute driven applications and services in research and science.

The main features are:

• ***Elastic computing infrastructure***. Execution of compute and data intensive applications (interactive and batch), hosting of long-time running services (web, application and database servers). Applications and services are hosted in the form of well fitted virtual machines (VM), running under system container technology, and scaled across multiple providers of the EGI federation.

• ***Virtual Machine (VM) image sharing and distribution***. Community-curated VMs and VM appliances are securely and automatically replicated across the infrastructure. EGI provides generic, baseline VM images, while user communities can offer specialised VMs and applications.

• ***Unified view of federation.*** EGI offers community federated accounting with an integrated view of the resource and service usage, abstracting the complexities of PaaS + SaaS to the final users.

The services provided by EGI to their 'Virtual Organisations' (VOs), are:

• **HTC Computing Services**, offered via HTC (high-throughput computing) clusters, shaped in IaaS clouds that support Docker containers to run computing or data intensive tasks and to host online services across institutional and national boundaries.

• **Storage services** enable scientific communities to store and share data within their reserved community space, and also shared with the broader public.

• **Tools, protocols and processes** to manage operational tasks of the infrastructure.

• **Training** about EGI services, IT Service Management (FitSM), and about the ISO27000 information security standard.

EGI groups scientific communities in 'Virtual Organisations' (VOs). A Virtual Organisation (VO) is usually composed of members who work and share the EGI resources in related research areas, or are part of the same scientific project. The members of a VO, use in EGI the same services, applications, data, and computing and storing capabilities.

Some of the most relevant success cases of VOs in EGI are ALICE, ATLAS, CMS and LHCb (Large Hadron Collider), the VIRGO experiment, the Cherenkov Telescope Array Observatory, and biomed, between others not so big communities, but relevant as international representation and research results.

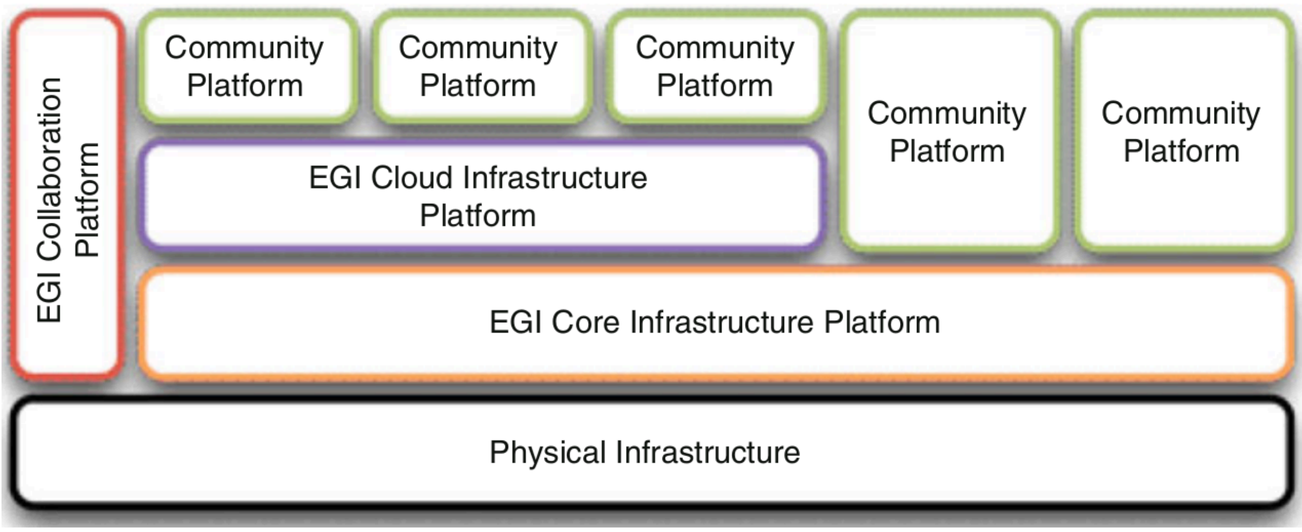
Since its start in 2010, EGI has had well defined processes to allocate resources (i.e. create and operate VOs) for large, structured, international user communities. These communities have a well established and long-term presence, and they are resourced well enough to sustain skilled IT support teams who can instantiate and operate VO services for the researchers. The most advanced research infrastructures from the ESFRI Research Infrastructure Roadmap are the typical operators of the largest EGI VOs.

3.7.1.2. EGI Architecture

The EGI architecture is organised in platforms (Figure 12):

• Core Infrastructure Platform, to operate and manage a distributed infrastructure.
• Cloud Infrastructure Platform, to operate a federated cloud-based infrastructure.
• Open Data platform, to provide easy access to large and distributed datasets.
• Collaboration Platform, for information exchange and community coordination.
• Community Platform.

Figure 12: EGI platform architecture [11]



Core Infrastructure Platform

The EGI Core Infrastructure provides all the necessary operational tools and processes to operate and manage a large distributed infrastructure guaranteeing standard operation of heterogeneous infrastructures from multiple independent providers.

This also includes:

• The Authentication and Authorisation infrastructure for homogeneous authentication and authorisation across the whole federation.

• A Service registry for configuration management of federated services.

• Monitoring tools, performing service availability monitoring and reporting of the distributed service end-points.

• Accounting for collecting, and displaying usage information.

• Information discovery about capabilities and services available in the federation.

• Virtual Machine image catalogue and distribution: allows researchers to share their virtual appliances for deployment in a cloud federation.

The federated environment of European e-Infrastructures, implemented in EGI through the Core Infrastructure platform, is a key enabler for distributed management and processing of big data and a fundamental baseline to implement the service integration and management system of the EOSC.

Collaboration platform

It provides IT Infrastructure and services that facilitate collaboration between research communities. Its two main components are the *Marketplace* and the *Application Database*.

The Marketplace has the ambition of becoming the platform where an ecosystem of EGI-related services, delivered by providers and partners, can be promoted, discovered, ordered, shared and accessed, including EGI offered services as well as discipline and community-specific tools and services enabled by EGI and/or provided by third parties under defined agreements.

The need of a Marketplace, making discoverable open research data and the related tools and knowledge, and acting as a one stop-shop for EU researchers, has been identified in the Service Hub model to build the EOSC and in the e-Infrastructure Reflection Group (e-IRG) roadmap [11]. This should also act as a single point of access to all e-Infrastructure services and tools for all users. The EGI Marketplace could be seen as the first test to implement such a tool (to be extended both in terms of features and coverage).

The Application Database (AppDB) is a tool that stores and provides information about:

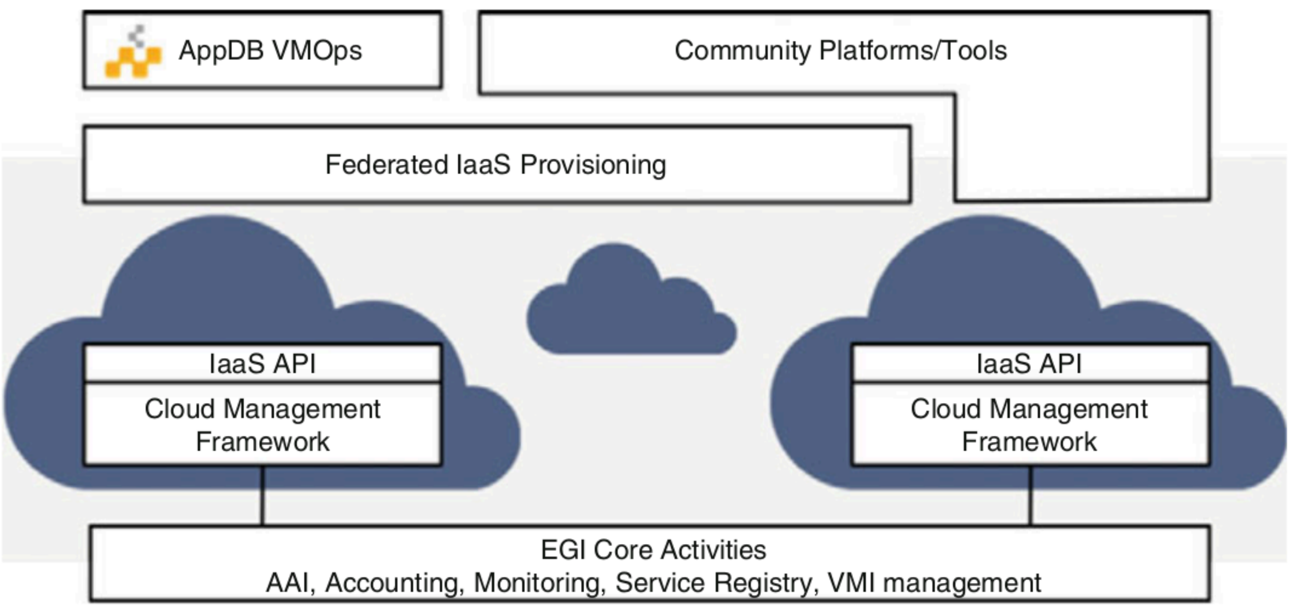
• Software solutions in the form of native software products and virtual appliances,
• The programmers and the scientists who are involved, and
• Publications derived from the registered solutions.

Reusing software products registered in the AppDB, means that scientists and developers may find a solution that can be directly utilised on the infrastructure. In this way, scientists can spend less or even no time on developing and porting a software solution to the Distributed Computing Infrastructures (DCIs) and facilitate the reproducibility of experiments. AppDB, thus, aims to avoid duplication of effort across the DCI communities and to inspire scientists who are less familiar with DCI programming and usage. The service is open to every scientist interested in publishing and therefore sharing their software solution.

Cloud Infrastructure platform

EGI launched the production phase of a cloud federation to serve research communities in May 2014, the EGI Federated Cloud (EGI 2017e). It integrates community, private and/or public clouds into a scalable computing platform for data and/or compute-driven applications and services. Its architecture is based on the concept of an abstract Cloud Management Framework (CMF) that supports a set of cloud interfaces to communities. Each resource centre of the infrastructure operates an instance of this CMF according to its own technology preferences and integrates it with the federation by interacting with the EGI Core Infrastructure platform. This integration is performed by using public interfaces of the supported CMFs, thus minimising the impact on site operations. Providers are organised into realms exposing homogeneous interfaces and grouping resources dedicated to serve specific communities and/or platforms. The EGI Federated Cloud is based on a hybrid model where private, community and public clouds can be integrated and already offers some of the facilities that a Service Hub should provide such as the virtualisation and the easy share and reuse of tools (Figure 13).

Figure 13: EGI federated cloud architecture



Each resource centre of the infrastructure operates an instance of a CMF according to its own technology preferences and integrates it with the federation by interacting with the EGI Core Infrastructure platform. Providers are organised in realms exposing homogeneous interfaces (Federated IaaS provisioning). Community platforms can exploit resources from one or more realms through such interfaces. The AppDB VMOps enables an automatic deployment of virtual appliances on all the resource centres supporting a specific community.

3.7.1.3. EGI Solutions applied to support JCORE Services

JCORE		EGI Solution
Integration Layer	Catalogue of Services	
Community	JWP - JCORE Web Portals	EGI Operations Portal
	VRE – Virtual Research Environment	EGI Notebooks
	TTC – Thematic & Technical Center	EGI Virtual Organisations
	RPS – Resources Providers	EGI Providers Services
	TNA – Transnational Access	EGI Configuration Database
	JCA – JCore API	EGI Configuration Database
	SCN – SoCial Networking	EGI Collaboration Tools
Process	HRV – HarVesting	EGI Configuration Database EGI Data Services EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
	HYD – HYDra	EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
	EVT - EvenT Engine	EGI Collaboration Tools EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
Information	VAMS – Virtual Access Metric System	EGI Accounting EGI Dynamic On-Demand Analysis Software (DODAS) EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
	DAA - DAta Analysis	EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
	DAP - DAta Process	EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)

	CTR - CaTalog of Resources	EGI Configuration Database + EGI Cloud Container Compute (ad hoc development needed)
Governance & System Management	AAI – Access & Auth. Infrastructure	EGI Check-in Service EGI Security Coordination EGI Configuration Database
	SYS – SYStem management	Infrastructure Manager Infrastructure Manager dashboard Infrastructure command-line interface EGI Elastic Cloud Compute Cluster EGI Workload Manager Dynamic On-Demand Analysis Software (DODAS)
	GOV - GOVernance	EGI Configuration Database EGI Accounting EGI Helpdesk
Application Platform	DBaaS - DataBase as a Service	EGI Data Services
	IDaaS - IDentification as a Service	EGI Check-in Service
	PaaS - Platform as a Service	EGI Applications Database
	IaaS - Infrastructure as a Service	EGI Federated Cloud Compute EGI Cloud Container Compute

3.7.2. EUDAT

3.7.2.1 Introduction

The EUDAT project is a pan-European data initiative that started in October 2011. The project brings together a unique consortium of 25 partners – including research communities, national data and high-performance computing (HPC) centres, technology providers, and funding agencies – from 13 countries. EUDAT aims to build a Collaborative Data Infrastructure (CDI) in a sustainable cross-disciplinary and cross-national data infrastructure that provides a set of shared services for accessing and preserving research data [13].

The EUDAT project was the first of a series of two funded projects aimed at building the foundation for the current Collaborative Data Infrastructure (CDI). The project has established the basis for the EUDAT CDI production infrastructure with a strong service-oriented approach and close relationship to research communities. It has managed to launch a set of well-identified services available to European researchers and research communities.

EUDAT2020 brought together a unique consortium of e-infrastructure providers, research infrastructure operators, and researchers from a wide range of scientific disciplines under several of the ESFRI themes, working together to address the new data challenge. In most research communities, there is a growing awareness that the "rising tide of data" will require new approaches to data management and that data preservation, access and sharing should be supported in a much better way. Data, and Big Data, is a cross-cutting issue touching all research infrastructures. EUDAT2020's vision was to enable European researchers and practitioners from any research discipline to preserve, find, access, and process data in a trusted environment, as part of a Collaborative Data Infrastructure (CDI) conceived as a network of collaborating, cooperating centres, combining the richness of numerous community-specific data repositories with the permanence and persistence of some of Europe's largest scientific data centres.

EUDAT2020 built on the foundations laid by the first EUDAT project, strengthening the links between the CDI and expanding its functionalities and remit. Covering both access and deposit, from informal data sharing to long-term archiving, and addressing identification, discoverability and computability of both long-tail and big data, EUDAT2020's services addressed the full lifecycle of research data.

One of the main ambitions of EUDAT2020 was to bridge the gap between research infrastructures and e-Infrastructures through an active engagement strategy, using the communities that were in the consortium as EUDAT beacons and integrating others through innovative partnerships.

During its three-year funded life, EUDAT2020 evolved the CDI into a healthy and vibrant data-infrastructure for Europe, and positioned EUDAT as a sustainable infrastructure within which the future, changing requirements of a wide range of research communities are addressed.

In February 2016, EUDAT signed a joint statement with LIBER (Association of European Research Libraries www.libereurope.eu), OpenAIRE (Open Access Infrastructure for Research in Europe www.openaire.eu), EGI (International Grid & Cloud Infrastructure for Research www.egi.eu) and GÉANT (European research and education networking collaboration www.geant.org) illustrating the shared vision of these organisations for the European Open Science Cloud (EOSC) for Research.

In January 2018, the EUDAT CDI has started working with a consortium of over 74 beneficiaries including Research Infrastructures, national e-Infrastructure providers, SMEs and academic institutions within the EOSC-Hub project, a Research and Innovation Action funded by the European Commission (call H2020- EINFRA-12-2017(a). EOSC-hub creates the integration and management system (the Hub) of the future European Open Science Cloud and the EUDAT CDI is a key part of it (www.eosc-hub.eu).

3.7.2.2 EUDAT Service Infrastructure

EUDAT's vision is that data is shared and preserved across borders and disciplines. Achieving this vision means enabling data stewardship within and between European research communities through a Collaborative Data Infrastructure (CDI), a common model and service infrastructure for managing data spanning all European research data centres and community data repositories [13].

Although research communities from different disciplines have different ambitions and approaches – particularly with respect to data organisation and content – they also share many basic service requirements. This commonality makes it possible for EUDAT to establish common data services, designed to support multiple research communities, as part of this CDI.

Since the EUDAT project has been constituted, it received the approaches and requirements of a subset of communities from linguistics (CLARIN), solid earth sciences (EPOS), climate sciences (ENES), environmental sciences (LIFEWATCH), and biological and medical sciences (VPH), regarding the deployment and use of a cross-disciplinary and persistent data e-Infrastructure. This analysis was conducted through interviews and frequent interactions with representatives of the communities. After several months of discussion and interaction with representatives from these communities, the basic and generic services have been identified by these communities as priorities (Figure 14). The services are: data replication from site to site, data staging to computer facilities, metadata, and easy storage. A number of enabling services, such as distributed authentication and authorization, persistent identifiers, hosting of services, workspaces and centre registry, were also discussed. The actual EUDAT's service catalogue is a set of seven major products (Figure 15) .

Figure 14: EUDAT CDI - Research Data Life Cycle

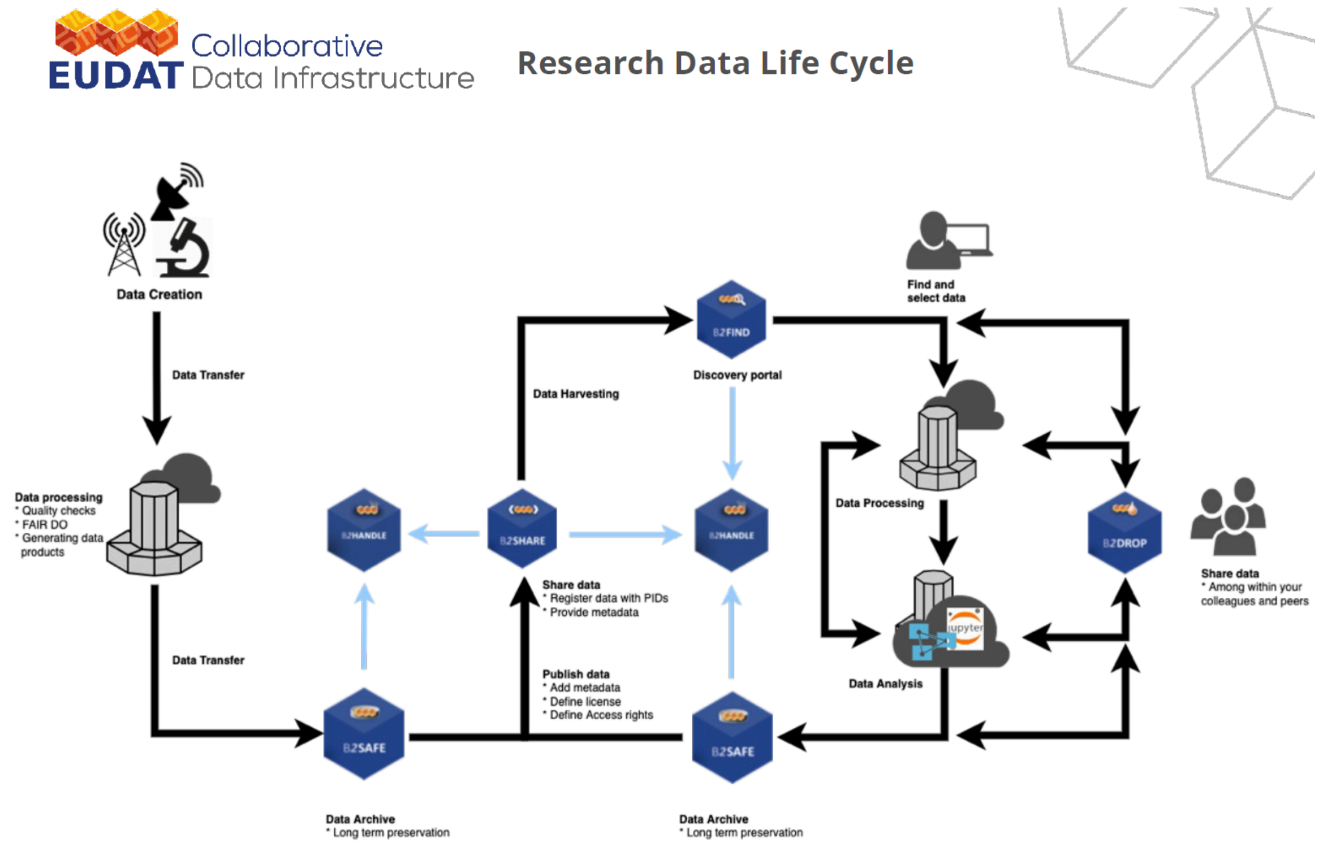
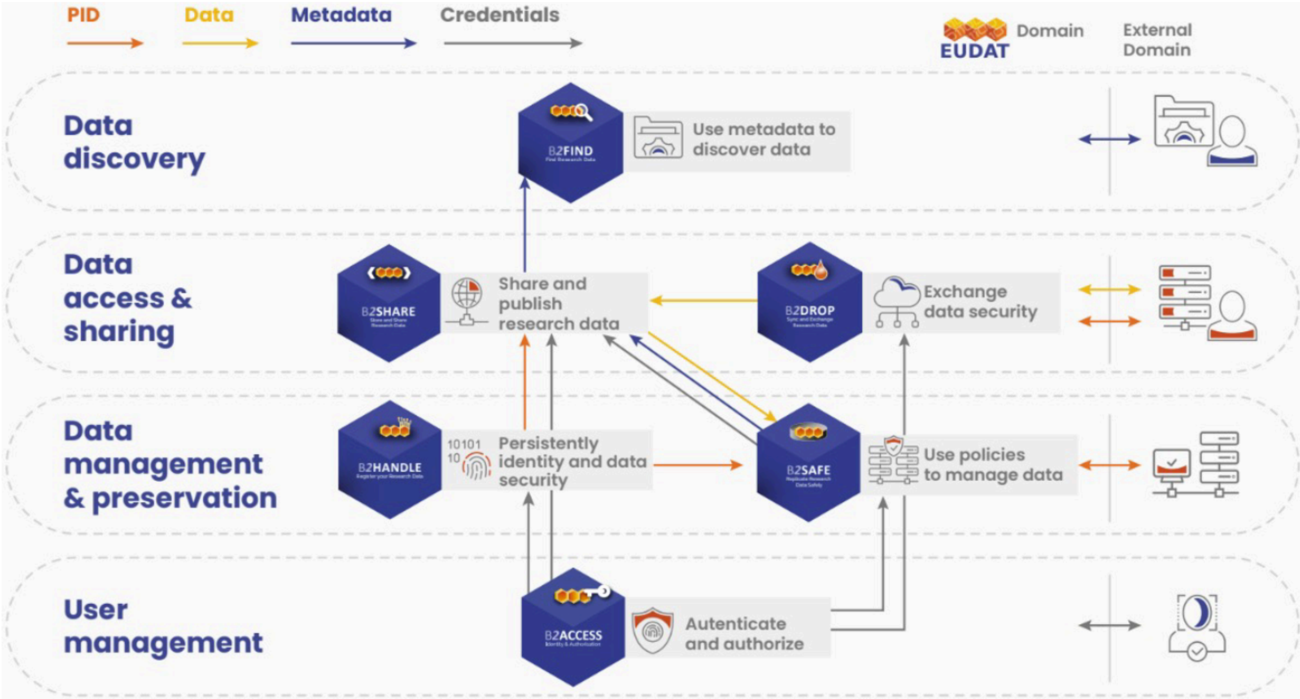


Figure 15: EUDAT Service Catalogue



- **B2FIND (**https://b2find.eudat.eu/**)** Is an interdisciplinary discovery portal for research output that allows free term search, results may be narrowed down using several facets, including spatial and temporal search options. B2FIND is also the EUDAT metadata indexing service and provides a discovery which allows users to find data collections within an international and inter-disciplinary scope. It is based on a comprehensive metadata catalogue of research data collections stored in EUDAT data centres and community repositories. Harmonisation of the metadata descriptions collected from heterogeneous sources enables not only the presentation in a consistent form but also the faceted search across scientific domain boundaries. For Communities and other providers of research data who need to publish and give visibility to their metadata and individual researchers who need to search data from everywhere, and see data in the context with an across community approach.

  Features:

  - Harmonisation of the metadata descriptions via the EUDAT Core metadata schema.
  - Harvesting of repositories via different protocols (e.g. OAI-PMH, CSW, Rest-APIs).
  - Faceted search via 17 facets (including geospatial, temporal search options), additional free text search.
  - Metadata aggregation from community repositories, multiple metadata standards are supported.
  - Harvested by OpenAIRE explorer.

- **B2SAFE** Is a robust and highly available service which allows community and departmental repositories to implement data management policies on their research data across multiple administrative domains in a trustworthy manner. It offers an abstraction layer of large scale, heterogeneous data storages, guards against data loss in long-term archiving, allows optimised access for users (e.g. from different regions), and brings data closer to facilities for compute-intensive analysis.

  Features:

  - Support for data management policies (e.g. registration of PIDs, cross-site replication, data integrity checks).
  - Support for policies customised to community and organisational needs.
  - Support for less frequently used archival data, but can also support active data.
  - Support for large scale storage resources (e.g up to PB-scale).
  - A single namespace across heterogeneous storages.
  - Supports integration with different kinds of storage systems (e.g. Tape based HSM, POSIX filesystems, Object storage).
  - Access via GridFTP, Webdav, iRODS commands.
  - Service offered by a network of EUDAT service providers-

- **B2SHARE (**https://b2share.eudat.eu/**)** Is a user-friendly, reliable and trustworthy way for researchers, scientific communities and citizen scientists to store, publish and share research data in a FAIR way. B2SHARE is a solution that facilitates research data storage, guarantees long-term persistence of data and allows data, results or ideas to be shared worldwide. B2SHARE supports community domains with metadata extensions, access rules and publishing workflows. EUDAT offers communities and organisations customised instances and/or access to repositories supporting large datasets.

  Features:

  - Support of metadata descriptions via the EUDAT metadata schema.
  - Registers DOIs for datasets and handles PIDs for data objects.
  - Supports versioning.
  - Harvested by B2FIND and OpenAIRE explorer.
  - Direct upload from B2DROP.
  - Accessible via Web GUI and API to support automatic publishing workflows.
  - Supports community domains.
  - Allows communities to define metadata extensions, access rules and publishing workflows.

- **B2DROP (**https://b2drop.eudat.eu/**)** Is a low-barrier, user-friendly and trustworthy storage environment which allows users to synchronise their active data across different desktops and to easily share this data with peers. EUDAT offers a free public basic instance for any researcher. For communities and organisations a premium service is offered on the public instance. Communities and organisations can also request customised instances.

  Features:

  - Default quota of 20GBs (Basic), high quotas optional (Premium).
  - Access via Web GUI, desktop clients and Webdav.
  - Multiple versions of files are kept.

- Enabled apps: Contacts, Calendar, Tasks, Circles (social communities).
- Sharing within B2DROP, across different instances (via API) and via links.
- Publishing of datasets to B2SHARE.
- Integration with CLARIN Language Resource Switchboard (Basic).
- Integration with other community services optional (Premium).
- Group management (Premium).
- OnlyOffice (Premium).

- **B2NOTE** Allows to easily create, search and manage annotations. An annotation is a keyword or commentary attached to a data object (data collection, file) that explains or classifies it. B2NOTE is a central service for annotating data content hosted within the EUDAT CDI.

  B2NOTE supports 3 types of annotations:

  - The semantic tag, a keyword from an ontology (a semantic tag coming from identified ontology repositories - currently only Bioportal),
  - the free-text keyword, to be created and used when a specific semantic term is not found,
  - the comment, a more comprehensive annotation.

  Relevant for:

  - Users must register to create and maintain annotations, additional access registration can be applied on the service which enables annotation functionality via B2NOTE.
  - Service providers willing to enable annotation via B2NOTE on an existing service can request consultancy and support.

  Features:

  - Is integrated into B2SHARE: access data via B2SHARE, then annotate them with B2NOTE.
  - Annotations are created and stored in a computer-readable format using the W3C Web Annotation model.
  - Annotations are searchable (in contrast to hand-written annotations).
  - Search results can be visualised in the User Interface and refined before being exported either as JSON-LD for reuse with your analysis workflow/script/software.
  - Allows users to export all the annotations about a file as JSON-LD for your own purpose.
  - This version is a crowdsourcing annotation service meaning that all annotations will be publicly available. To preserve the anonymity as a user you can register and create an annotator pseudonym.

- **B2ACCESS (https://b2access.eudat.eu/)** Is a federated cross-infrastructure authorisation and authentication proxy for user identification and community-defined access control enforcement. It allows users to authenticate themselves using a variety of credentials providing federated access and single-sign-on to services and service providers in a trusted way. B2ACCESS offers communities and service providers an AARC compliant AAI proxy ready to be integrated within the EOSC AAI Federation.

  Features

  - Compliant to the AARC Blueprint and REFEDS Sirtfi.
  - Supports group management.
  - Supports authorisation via group membership.

- Supports authentication e.g. via eduGAIN, ORCID and Social Identities
- Support for EUDAT B2ACCESS local accounts
- Support for IdP and service integration via SAML, OAuth and OIDC

- **B2HANDLE** Is the distributed service for storing, managing and accessing persistent identifiers (PIDs) and essential metadata (PID records) as well as managing PID namespaces. The implementation of the service relies on the DONA/Handle persistent identifier solution. B2HANDLE can be used by middleware applications, end-user tools and other services to reliably identify data objects over longer time spans and through changes in object location or ownership. The B2HANDLE service encompasses management of identifier namespaces (handle prefixes), establishment of policies and business workflows, operation of handle servers and technical services, and a user-friendly Python library for general interaction with handle servers and EUDAT-specific extensions. B2HANDLE is mostly transparent to end-users, shielding them from the complexity of infrastructure details. B2HANDLE supports a dedicated Handle record structure (a PID profile) for the safe data management within an infrastructure with a given topology.

  Features:

  - Globally resolvable identifiers via the Global Handle Network (DONA, https://hdl.handle.net/).
  - Communities and organisations can obtain Handle Prefixes (ePIC spec.) for their own use.
  - PIDs can be hosted at EUDAT service providers.
  - B2HANDLE operates as a federation of EUDAT service providers based on policies.
  - PIDs are mirrored across multiple providers for high resolution and high availability.
  - Handle records can be customised to community and organisational needs (for example include checksums and timestamps to ensure authenticity of the data objects).
  - Support for reverse look-ups of Handle records.
  - REST API for easy registration and minting of PIDs.

3.7.2.3 EUDAT Solutions applied to support JCORE Services

JCORE		EUDAT Solution
Integration Layer	Catalogue of Services	
Community	JWP - JCORE Web Portals	B2FIND, B2SHARE, B2DROP
	VRE – Virtual Research Environment	There is no specific VRE available in EUDAT. It should be implemented a part (i.e. Jupyter Notebook server), but all information and resources in EUDAT (via APIs) are accessible through B2FIND, B2SAFE, B2SHARE, B2DROP (with Webdav protocol), and secured with B2ACCESS.
	TTC – Thematic & Technical Center	No specific solution available in EUDAT. The backend could be implemented using the APIs of B2FIND, B2SAFE, B2SHARE (deeper study needed). The frontend should be developed out of EUDAT.
	RPS – Resources Providers	Could be developed as a client (out of EUDAT), using B2ACCESS, B2SHARE, B2HANDLE, B2SAFE APIs
	TNA – Transnational Access	Not available in EUDAT. Should be developed.
	JCA – JCore API	Not available in EUDAT. Should be developed.
	SCN – SoCial Networking	No specific solution available in EUDAT. The backend could be implemented using the APIs of B2FIND, B2SAFE, B2SHARE (deeper study needed). The frontend should be developed out of EUDAT, using B2ACCESS as authentication and authorization mechanism.
Process	HRV – HarVesting	Could be solved as the case of B2FIND and OpenAIRE harvesting combination works. Deeper study is needed.
	HYD – HYDra	Not available in EUDAT. Should be developed OUT OF EUDAT (not clear how to integrate custom developments)
	EVT - EvenT Engine	Not available in EUDAT. Should be developed OUT OF EUDAT (not clear how to integrate custom developments)
Information	VAMS – Virtual Access Metric System	Not available in EUDAT. Should be developed OUT OF EUDAT (not clear how to integrate custom developments)
	DAA - DAta Analysis	Not available in EUDAT. Should be developed OUT OF EUDAT (not clear how to integrate custom developments)
	DAP - DAta Process	Not available in EUDAT. Should be developed OUT OF EUDAT (not clear how to integrate custom developments)

	CTR - CaTalog of Resources	EUDAT contains its own database, ready to be used via B2FIND, B2SHARE, B2DROP but the internal structure belongs to EUDAT and it is not modifiable for custom purposes.
Governance & System Management	AAI – Access & Auth. Infrastructure	B2ACCESS.
	SYS – SYStem management	Not available but, due the completeness of EUDAT SLAs, is not needed. It comes as an external service.
	GOV - GOVernance	B2ACCESS, B2SAFE + EUDAT SLAs
Application Platform	DBaaS - DataBase as a Service	Performed by EUDAT SLAs.
	IDaaS - IDentification as a Service	B2ACCESS under EOSC Hub project, leverages a complete AAI platform.
	PaaS - Platform as a Service	Performed by EUDAT SLAs.
	IaaS - Infrastructure as a Service	Performed by EUDAT SLAs.

3.7.3. EOSC

3.7.3.1 Introduction

As part of the European Commission Digital Single Market strategy (European Commission 2015 [15]), the European Open Science Cloud (EOSC) initiative was officially launched in April 2016 by the European Commission. EOSC promotes not only scientific excellence and data reuse but also job growth and increased competitiveness in Europe, and drives Europe-wide cost efficiencies in scientific infrastructures through the promotion of interoperability on an unprecedented scale. [11]

According to the first report of the High Level Expert Group on the European Open Science Cloud (EOSC) (European Commission 2016 [16]) appointed by the European Commission, EOSC has been defined as a support environment for Open Science aiming to "*accelerate the transition to more effective Open Science and Open Innovation in a Digital Single Market by removing the technical, legislative and human barriers to the re-use of research data and tools, and by supporting access to services, systems and the flow of data across disciplinary, social and geographical borders*". Indeed, the term "cloud" has been interpreted as "a metaphor to help convey the idea of seamlessness and a commons".

As guiding principles, the experts underlined how:

1. The EOSC *must integrate with other e-Infrastructures and initiatives in the world*, implementing a lightweight interconnected system of services and data, which follows the federated model.

2. Open refers to the accessibility of services and data under proper non-discriminatory policies ("*not all data and tools can be open*" and "*free data and services do not exist*").

3. EOSC should include all scientific disciplines.

4. The term cloud ***should not*** refer to ICT infrastructure but to a commons of data, software, standards, expertise and a policy framework relevant to data-driven science and innovation.

3.7.3.2 Minimum Viable & Future EOSC Architecture

In December 2019 the Sustainability WG (ESWG) released the Tinman report on Solutions for a Sustainable EOSC [17]. Via an open consultation process this report has evolved into the FAIR Lady report [18] which was published on the 24th of November 2020. The scope of the FAIR Lady report is to explore possible means for sustaining the European Open Science Cloud beyond the initial phase which terminates at the end of 2020. To better understand the elements that are required to be sustained, the ESWG described in the FAIR Lady document a first iteration of a Minimum Viable EOSC (MVE). On the basis of the FAIR Lady report, the EAWG has developed a high-level view (see Figure 16) to place the EOSC-Core, EOSC-Exchange and MVE in context with the Research Infrastructure and e-Infrastructures (i.e., EOSC Federation) and the research community at large.

Figure 16: High-level view of MVE - Minimum Viable EOSC Architecture

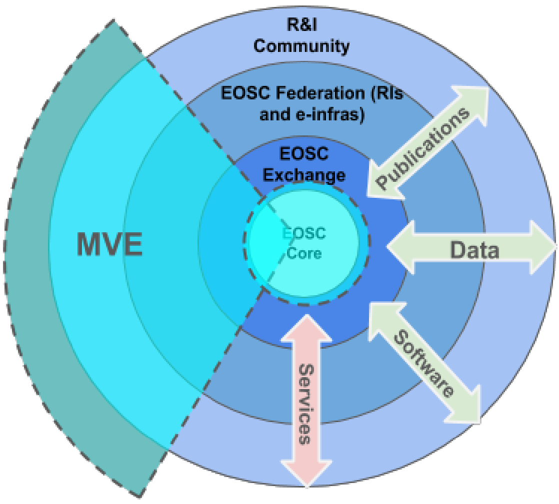
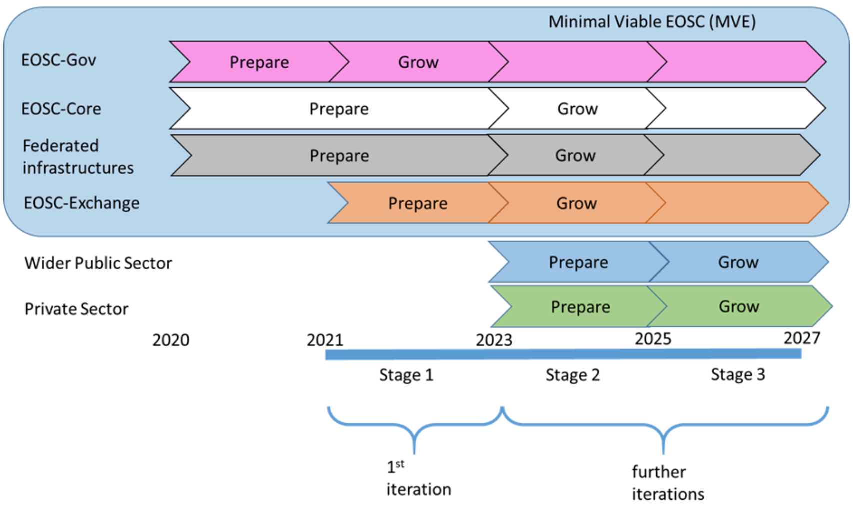


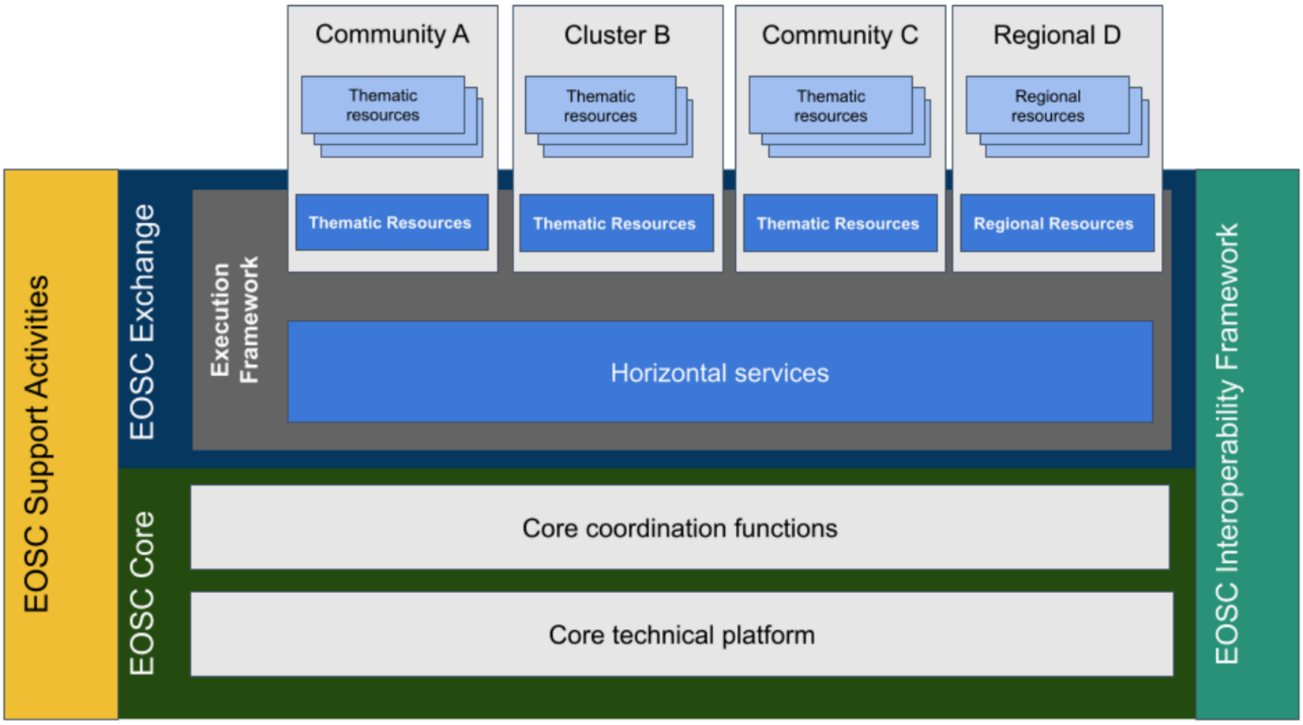
Figure 17 shows the timeline for the EOSC MVE (more detailed at the EOSC Strategic Research and Innovation Agenda (SRIA) [19][20].

Figure 17: Timeline of the MVE - Minimum Viable EOSC Architecture



Actually, the future EOSC architecture is being defined on the output of the EOSC Interoperability Framework Report [21][22][23], the past EOSC projects (EOSC Enhance, EOSC-hub, OpenAIRE Advance, etc.), and on work carried out during the EOSC Future project preparation. This involves not only defining the 'layers' of EOSC but also the elements and the glue within them and how they are interconnected. A high- level view of the EOSC architecture is shown in Figure 18.

Figure 18: High-level view of the (future) EOSC Architecture



The future EOSC Architecture elements:

EOSC-Core. Is the set of internal services which allow EOSC to operate. It includes a core technical platform that facilitates EOSC operations, upon which the researcher-facing resources in the EOSCExchange can rely and integrate as appropriate. It also includes non-technical coordination functions, such as the onboarding and security coordination, which operate and facilitate the technical platform.

EOSC-Exchange. Is the set of federation services and other resources registered into the EOSC by Research Infrastructures and Science Clusters to serve the needs of research communities. Generic services and resources which target heterogeneous scientific domains and research communities are identified as *Horizontal Services*. Resources that target users from a specific science, community and/or regional domain are identified as Thematic and/or Regional Resources. The capability to compose resources across horizontal and thematic and/or regional resources in compliance with the EOSC Interoperability Framework is defined as the *Execution Framework*. While it is expected that the majority of the Horizontal Services are provided by the eInfrastructures (e.g., EGI, EUDAT, OpenAIRE, GÉANT), generic services and resources offered by the Science Cluster communities will also be offered as a horizontal service.

EOSC Interoperability Framework (EOSC IF). Provides the procedures and services required to support a flexible framework of standards and guidelines that facilitate the interoperability and composability of EOSC resources in the EOSC-Exchange via the EOSC-Core. As such, it leverages a *semantic overlay* where EOSC resources can be associated to the standards and guidelines (the IFs) they comply to, and therefore, be related by composability and interoperability features, across communities and providers. The EOSC IF is defined as a Reference Architecture Framework, which enables and governs: - A 'System of Systems' EOSC-Core architecture, via consultation and consolidation with the communities: the set of interoperability guidelines required for EOSC Providers to engage with the EOSC-Core and benefits from its added value services; - A registry database of EOSC Interoperability Guidelines as defined, used, and proposed by the communities, which enables providers to clearly specify the interoperability boundary of EOSC resources and oversee the interoperability frameworks adopted by communities.

EOSC Support activities. Sit alongside the EOSC-Core and EOSC-Exchange, and comprise the training, engagement, and other human-centric activities which make EOSC more attractive and easier to use, and help users benefit from it more easily once engaged.

Science Clusters and Communities. Are embedded in EOSC through the work of EOSC Future but will continue to operate outside of the EOSC for their specific community. These include the Science Clusters (from the INFRAEOSC-04 call), the Regional Initiatives (from the INFRAEOSC-05 call), as well as national communities, other research communities and less organised groups from the long tail of science and research. They will bring a rich set of resources to EOSC but will also have resources and other elements outside EOSC, which are targeted to their own individual communities, including richer ontologies and domain-specific information and support.

3.7.3.3 EOSC System Architecture

The Deliverable D5.4 - Final EOSC Service Architecture[24] of the EOSCpilot -WP5, builds upon the foundations defined in the initial EOSC Service Architecture [25], and realises the MVE Architecture on the final EOSC IT System.

It is important to note that this architecture model expresses a conservative view of Open Science Services. It reflects practices and functionalities that are available today to support the Open Science approach. In this sense, it represents what is named a Minimal Viable Product [26], i.e. a product with just enough features to satisfy early customers, and to provide feedback for future product development. It is expected that in the near future many innovative functionalities and classes of services will emerge, as a consequence of a better understanding of new Open Science practices, and the expansion of the new market of third-party service providers.

EOSC System is planned to offer services belonging to the following classes of service (Figure 19):

- ***Front-end services***. Implements the part of the overall service with which users will interact directly, namely portals or APIs.

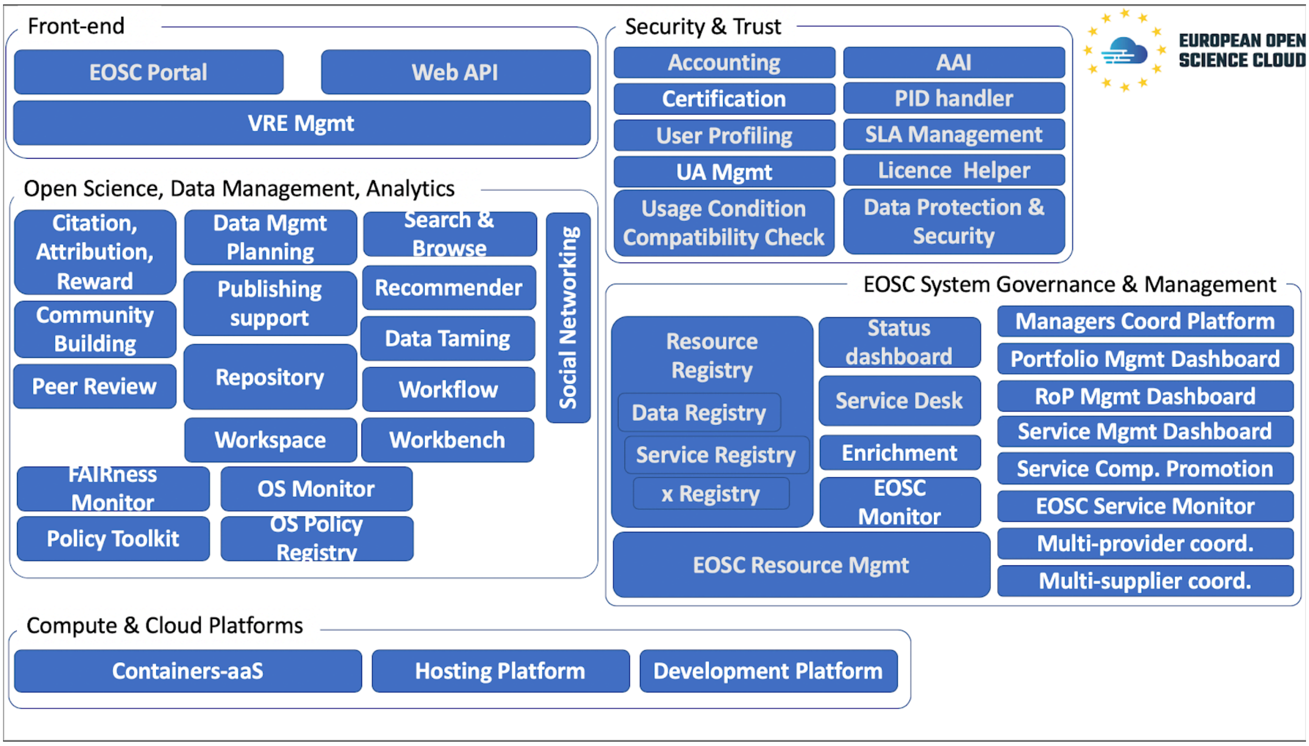
- ***Security & Trust.*** Aims to guaranteeing that the overall system (and the services) operate securely and according to standard.

- ***Open Science, Data Management, Analytics***. Provide their users with user and open science friendly facilities, enabling users to focus on science tasks.

- ***EOSC System Governance & Management***. Supports the operation and management of the overall EOSC System.

- **Compute & Cloud Platforms.** Offer generalist resources like virtual machines and containers as well as network transport connectivity. In addition, all the platforms and software that do not belong to the other categories fall here.

Figure 19: EOSC System Architecture (final version) [23]



3.7.3.4. EOSC Solutions applied to support JCORE Services

JCORE		EOSC Solution
Integration Layer	Catalogue of Services	
Community	JWP - JCORE Web Portals	[EOSC Portal]()
	VRE – Virtual Research Environment	VRE Mgmt
	TTC – Thematic & Technical Center	Community Building FAIRness Monitor Policy Toolkit Citation, Attribution, Reward Peer Review Search & Browse Data Taming
	RPS – Resources Providers	Resource Registry EOSC Resource Mgmt Multi-provider coord. Multi-supplier coord.
	TNA – Transnational Access	Community Building Data Mgmt Planning Repository Publishing Support Search & Browse Workflow
	JCA – JCore API	Web API
	SCN – SoCial Networking	Social Networking
Process	HRV – HarVesting	Containers-aaS Development Platform

	HYD – HYDra	Hosting Platform (Ad hoc development needed)
	EVT - EvenT Engine	
Information	VAMS – Virtual Access Metric System	
	DAA - DAta Analysis	
	DAP - DAta Process	
	CTR - CaTalog of Resources	Repository
Governance & System Management	AAI – Access & Auth. Infrastructure	Security & Trust Services
	SYS – SYStem management	EOSC System Governance & Management Services
	GOV - GOVernance	
Application Platform	DBaaS - DataBase as a Service	Compute & Cloud Platform Services
	IDaaS - IDentification as a Service	
	PaaS - Platform as a Service	
	IaaS - Infrastructure as a Service	

3.7.4. D4Science

3.7.4.1 Introduction

D4Science is an organisation operating a data infrastructure offering a rich array of services by community-driven virtual research environments, focused in support communities involved in open science practices.

The D4Science initiative has been developed and supported by several European projects. DILIGENT [27] (2004-2007) in the 6th Framework Programme for integrating and strengthening the European research area, was the forerunner of a test infrastructure built by integrating digital library and grid computing technologies and resources. D4Science was conceived and developed to serve the needs of communities of practices involved in knowledge development.

The development of the D4Science initiative started with the support of D4Science [28] (2008-2009), D4Science-II [29] (2009-2011), ENVRI [30] (2011-2014), EUBrazilOpenBio [31] (2011-2013), iMarine [32] (2011-2014). In this period the infrastructure was established and developed to serve communities of practices from domains ranging from Earth Science to Marine Science with worldwide scope.

The following projects contributed to continue the D4Science development: BlueBRIDGE [33] (2015-2018), EGIEngage [34] (2015-2017), ENVRIplus [35] (2015-2019), Parthenos [36] (2015-2019), SoBigData [37] (2015-2019), AGINFRAplus [38] (2017-2019), PerformFish [39] (2017-2022), ARIADNEplus [40] (2019-2022), EOSC-Pillar [41] (2019-2022), DESIRA [42] (2019-2023), RISIS2 [43] (2019-2022), Blue-Cloud [44] (2019-2022), SoBigData++ [45] (2020-2023).

Nowadays, D4Science hosts the marine-science EOSC side, the actually most important federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters: the Blue-Cloud Gateway.

3.7.4.1 D4Science System Architecture

D4Science provides researchers and practitioners with a working environment where open science practices are transparently promoted [46][47]. This infrastructure integrated in gLite [48], is capable to exploit the EGEE [49], a largest grid infrastructure

D4Science is built and operated by relying on gCube[50][51] technology, a software-framework system specifically conceived to enable the construction and development of Virtual Research Environments (VREs).

Each VRE is equipped with basic services supporting collaboration and cooperation among its users, namely:

- A shared workspace to store and organise any version of a research artefact;
- A social networking area to have discussions on any topic (including working version and released artefacts) and be informed on happenings;
- A data analytics platform to execute processing tasks either natively provided by VRE users or borrowed from other VREs to be applied to VRE users' cases and datasets;
- A catalogue-based publishing platform to make the existence of a certain artefact public and disseminated.

The VREs are defined and deployed to serve the needs of distributed Virtual Organisations. The user communities involved in D4Science may be represented by one or more VOs. To support the provision of such diverse environments the gLite and gCube nodes play a fundamental role:

• gCube nodes are the hardware resources able to run gCube services. The gCube software includes a special web service container, the gCube Hosting Node (gHN), and a set of services and libraries that provide the functionality to create, manage, and exploit VREs;

• gLite nodes are computing and storage nodes running gLite software. By running gLite, these nodes provide core grid functionalities such as file-based storage, distributed computation of applications, etc. gLite nodes are exploited by gCube services which then provide higher level functionality through the D4Science VREs. Figures 20, 21 show the D4Science System Reference Architecture and the gHN Reference Architecture

The D4Science infrastructure is built by sites coming from different organisations participating in the D4Science project, i.e., CNR, ESA, FAO, National Kapodistrian Univ. of Athens (NKUA), and Univ. of Basel (UNIBASEL). The number of sites involved and the number of nodes allocated has gradually grown over the time to meet the demand related to the provision of new VOs and VREs. These nodes are exploited to run the gCube system delivered by D4Science and the gLite middleware as released by the EGEE [49] project.

The management of the infrastructure is facilitated by the definition and implementation of clear procedures for monitoring and accounting. A number of monitoring tools have been deployed allowing different infrastructures roles (Site Managers, Data Managers, etc.) to visualise the status of their resources and to be actively notified when problems occurred. An accounting tool has also been put in production providing relevant statistics about the users' exploitation of the infrastructure, and the service-to-service communications at infrastructure level.

Figure 20: D4Science System Reference Architecture

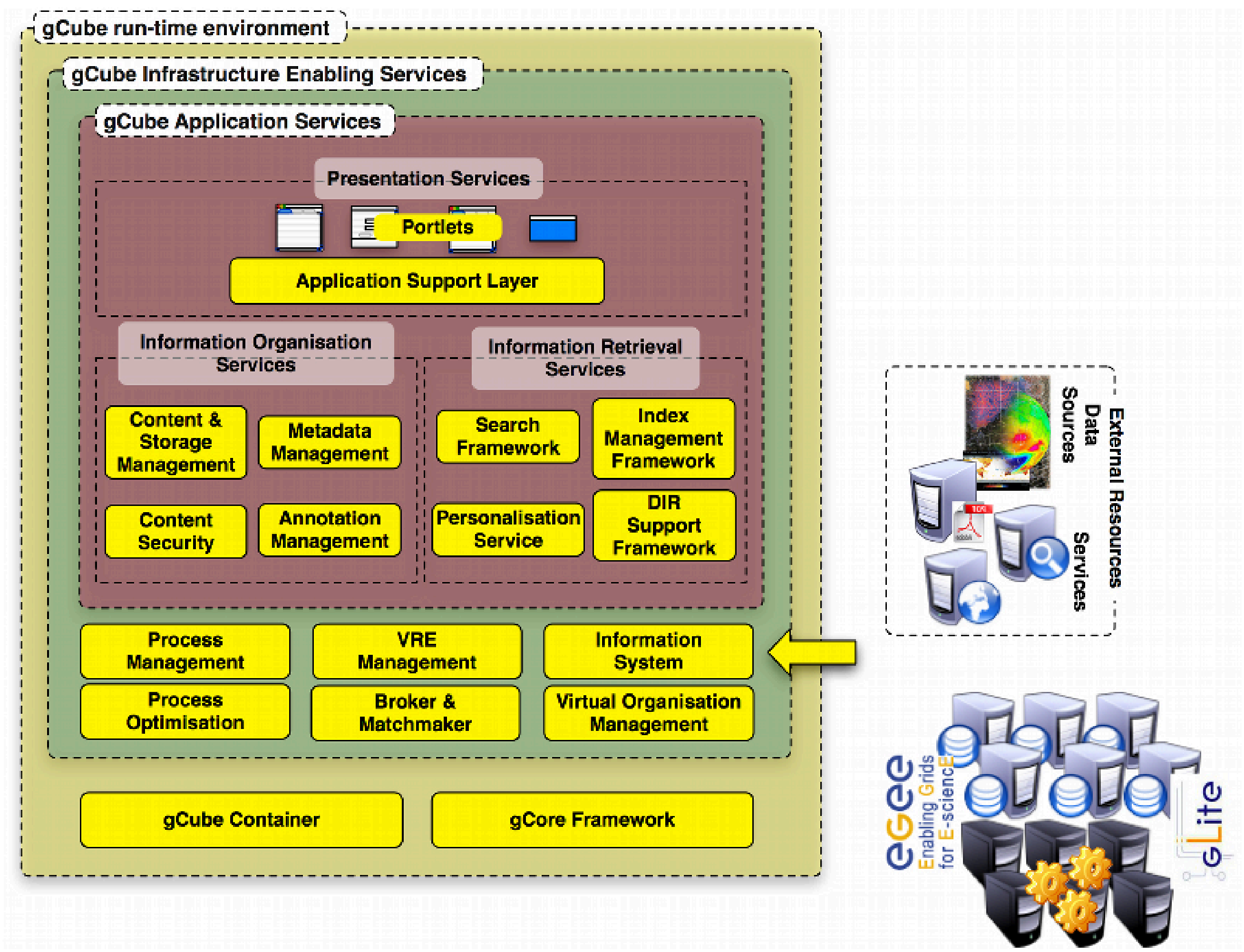


Figure 21: gHN Reference Architecture (runtime environment)

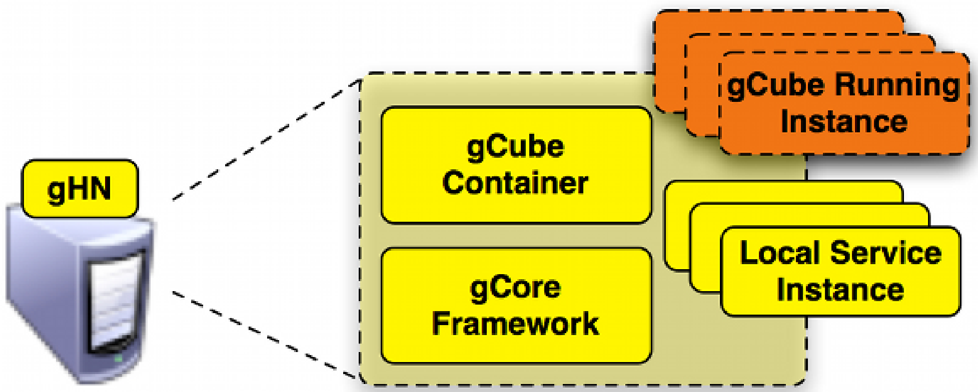
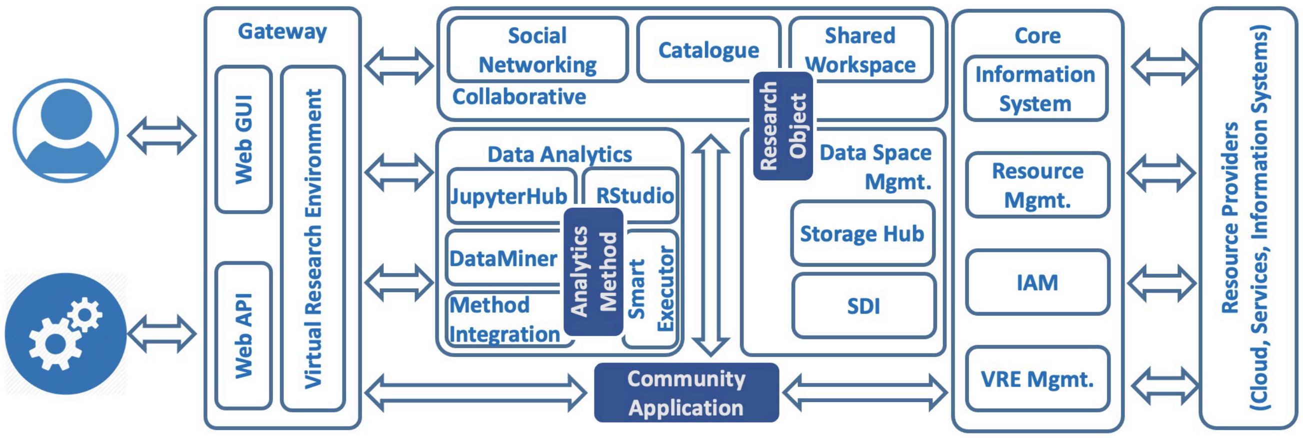


Figure 22: D4Science VRE - Overall Architecture



3.7.4.2. D4Science Solutions applied to support JCORE Services

JCORE		D4Science Solution
Integration Layer	Catalogue of Services	
Community	JWP - JCORE Web Portals	Gateway - Web GUI
	VRE – Virtual Research Environment	Gateway - Virtual Research Environment Core - VRE Mgmt.
	TTC – Thematic & Technical Center	Gateway - Virtual Labs, VRE Core - VRE Mgmt.
	RPS – Resources Providers	Resource Providers - Cloud, Services, Information Systems
	TNA – Transnational Access	Community Application - Gateway, Collaborative Resource Providers - Cloud, Services, Information Systems
	JCA – JCore API	Not available, but is possible to develop ad hoc software running in Docker.
	SCN – SoCial Networking	Collaborative - Social Networking
Process	HRV – HarVesting	Data Analytics - Data Miner Data Management Space
	HYD – HYDra	Not available, but is possible to develop ad hoc software running in Docker.
	EVT - EvenT Engine	Not available, but is possible to develop ad hoc software running in Docker.
Information	VAMS – Virtual Access Metric System	Not available, but is possible to develop ad hoc software running in Docker.
	DAA - DAta Analysis	Data Analytics
	DAP - DAta Process	Data Analytics

	CTR - CaTalog of Resources	Data Space Management
Governance & System Management	AAI – Access & Auth. Infrastructure	Core - IAM
	SYS – SYStem management	Core - Information System
	GOV - GOVernance	Core - IAM, Resource Management, VRE Management
Application Platform	DBaaS - DataBase as a Service	Performed by D4Science through SLA
	IDaaS - IDentification as a Service	Performed by D4Science through SLA
	PaaS - Platform as a Service	Performed by D4Science through SLA
	IaaS - Infrastructure as a Service	Performed by D4Science through SLA

3.7.5. Blue-Cloud

3.7.5.1 Introduction

The European Open Science Cloud (EOSC) was launched by the European Commission in 2016 to provide a virtual environment with open and seamless services for storage, management, analysis and re-use of research data, across borders and disciplines. Within this framework, Blue-Cloud, as the "Future of Seas and Oceans Flagship Initiative" of EU HORIZON 2020 programme, piloted the thematic EOSC for the marine domain, serving the Blue Economy, Marine Environment and Marine Knowledge agendas.

Blue-Cloud delivered a collaborative virtual environment to enhance FAIR and Open Science. Between October 2019 and March 2023, Blue-Cloud deployed a cyber platform with smart federation of an unprecedented wealth of multidisciplinary data repositories, analytical tools, and computing facilities to explore and demonstrate the potential of cloud-based Open Science and address ocean sustainability, UN Ocean Decade, and G7 Future of the Oceans objectives.

Blue-Cloud federated leading European marine Research infrastructures (SeaDataNet, EurOBIS, Euro-ARGO, ICOS, SOCAT, ENA, EMODnet, Copernicus Marine Service) and e-Infrastructures (EUDAT, D4Science, WEkEO DIAS), allowing researchers to combine, reuse, and share quality data across disciplines and countries. The federation takes place at the levels of data, computing and analytical service resources. [www.blue-cloud.org]

The actual Blue-Cloud 2026 has received funding from the European Union's Horizon Europe programme call HORIZON-INFRA-2022-EOSC-01, Project ID 101094227. Blue-Cloud has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n.862409 The views and opinions expressed in this website are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

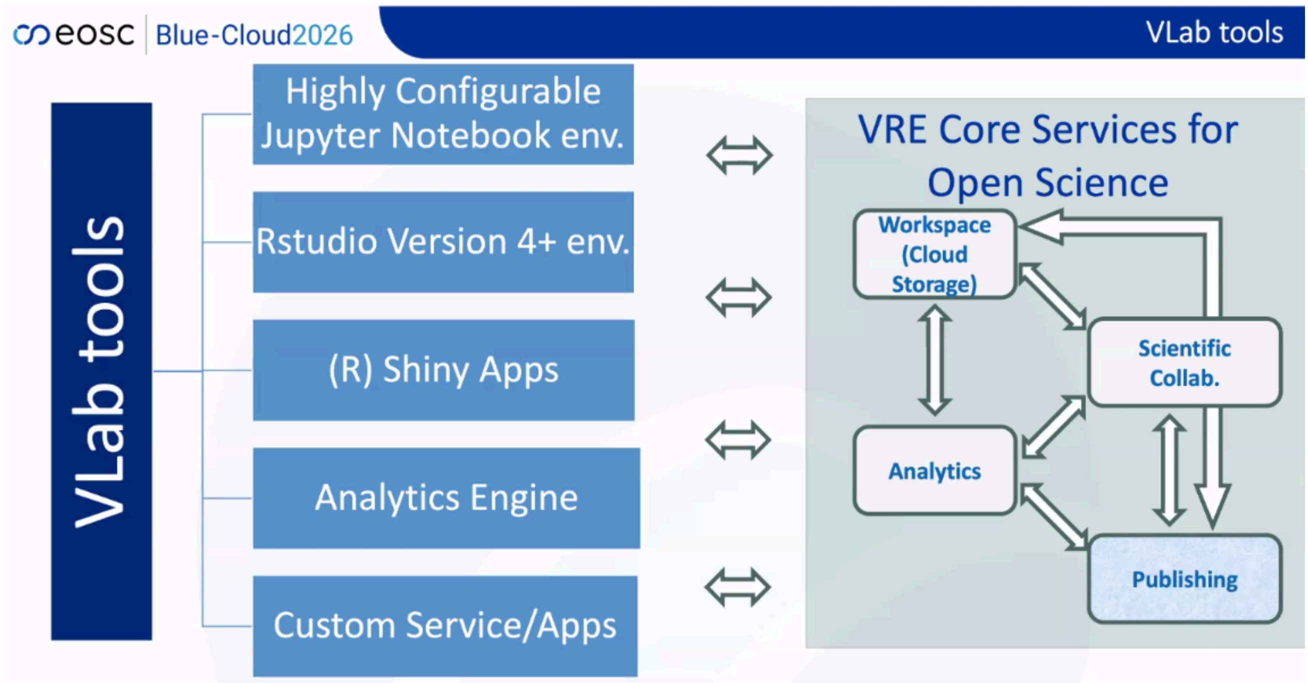
Blue-Cloud aims to build and demonstrate as a thematic EOSC cloud to support research for understanding & better managing the many aspects of ocean sustainability, ranging from sustainable fisheries to ecosystem health to pollution. The project seeks to bring together leading European marine data management infrastructures, EOSC horizontal e-infrastructures, and selected marine researchers to capitalise on what exists already and to develop and deploy the pilot Blue Cloud as a cyber platform bringing together and providing access to 1) multidisciplinary data from observations and models, 2) analytical tools, and 3) computing facilities that are essential for key blue science use cases. Those will demonstrate the potential of cloud based science, and also identify the longer term challenges for wider deployment in technical, organisational, and financial sense. Together these will provide the basis for the Blue-Cloud Strategic Roadmap, that might be implemented in the upcoming sustainable development goal #14 of the UN 2030 Agenda for Sustainable Development.

Since October 2019, the pilot Blue-Cloud project (www.blue-cloud.org) combined both the interests of the European Open Science Cloud (EOSC), aiming to provide a virtual environment with open and seamless access to services for storage, management, analysis and re-use of research data, across borders and disciplines, and the blue research communities by developing a collaborative web-based environment providing simplified access to an unprecedented wealth of o multi-disciplinary datasets from observations, analytical services, and computing facilities essential for blue science.

Blue-Cloud 2026 aims at a further evolution of this pilot ecosystem into a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for deepening research of oceans, EU seas, coastal & inland waters. It develops a thematic marine extension to EOSC for open web-based science, serving the needs of the EU Blue Economy, Marine Environment and Marine Knowledge agendas.

The Blue-Cloud VRE will be based upon the existing D4Science e-infrastructure as developed and managed by CNR-ISTI. This e-infrastructure already hosts multiple Virtual Labs [Figure 23] and offers a variety of services, which can be adopted for the Blue-Cloud. The D4Science e-infrastructure also has proven solutions for connecting to external computing platforms and means for orchestrating distributed services, which will be instrumental for smart connections to the other e-infrastructures in the Blue-Cloud system.

Figure 23: Blue-Cloud VLab tools.



The Blue-Cloud demonstrators will be developed as Virtual Labs embedded in the D4Science VRE e-infrastructure and supported by data input from the Blue-Cloud Data Discovery and Access Service (DD&AS) and other data resources, and possible additional computing services. The demonstrators are being worked on analysing their scientific workflows and technical set-up, and considering the present D4Science VRE infrastructure and services that will provide the basis platform.

The initial Blue-Cloud architecture as described in this report, is designed to be scalable and sustainable for near-future expansions, such as connecting additional blue data infrastructures, implementing more and advanced blue analytical services, configuring more dedicated Virtual Labs, and targeting more (groups of) users.

3.7.5.2 Blue-Cloud Architecture

The two main technical features of the Blue-Cloud Architecture are the:

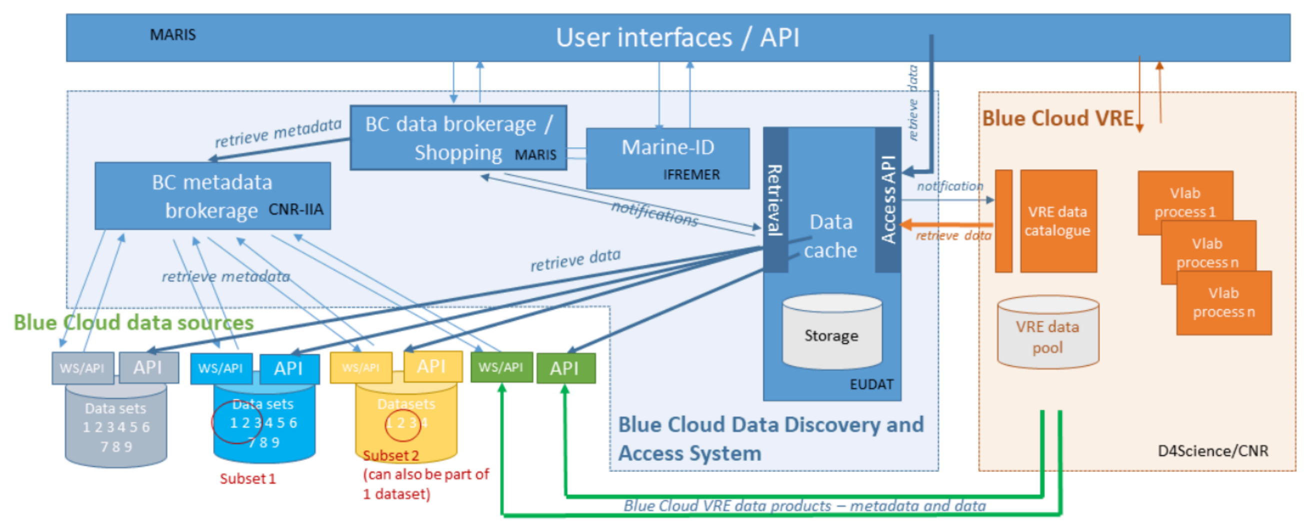
- Blue Cloud Virtual Research Environment (VRE)
- Blue Cloud Data Discovery and Access Service

As explained in the previous section, Blue-Cloud relies on the D4Science architecture, and the Blue Cloud VRE is in the DNA of D4Science. Keeping in mind that the D4Science VRE is a general purpose tool, what differentiates Blue-Cloud from other D4Science Gateways, is the *Blue Cloud Data Discovery and Access Service (DD&AS)*.

One of the biggest challenges in the Blue-Cloud project is federating resources, including data, computing, and analytical services. The Blue Cloud Data Discovery and Access service aims to unite existing blue data infrastructures (BDI) and allow multidisciplinary datasets to be shared with external users, serving as a primary source of data for the Blue Cloud VRE. Users will also be able to integrate external data sources, including their own, by utilising the internal VRE functionality for data ingestion and establishing direct access to select remote blue data infrastructures. This will allow a broad range of input data sources, such as in-situ data, earth observation data, and model outputs, to be utilised. By creating a federation through the Blue Cloud Virtual Research Environment (VRE), computing and analytical services can be shared and used in combination for specific applications. These analytical services can include a variety of algorithms and generic services for subsetting, pre-processing, publishing, and viewing data and data products.

The Blue Cloud Data Discovery and Access service architecture is based upon a combination of the DAB (Data Access Broker) metadata broker service of CNR-IIA, and the SeaDataNet CDI (Common Data Index) service modules as developed by MARIS, IFREMER, and EUDAT in the framework of the EU SeaDataCloud project. For the Blue-Cloud Data Discovery and Access service and its modules, additional developments were needed such as adapting and upgrading of existing services, adding new services, testing modules, integrating modules, and testing the integrated service, in order to achieve the planned functionality.[52]

Figure 24: Architecture of the Blue Cloud discovery and access service



The Blue Cloud Data Discovery and Access service consists of a number of modules (services) as indicated in the figure above [51] :

- **Blue Cloud metadata brokerage**, operated by CNR-IIA, dynamically interacts with each of the blue data infrastructures to retrieve, extract and harmonise metadata entries for each blue data infrastructure into a common Blue Cloud level 1 metadata catalogue.

- **Blue Cloud data sources**, comprising blue data infrastructures, that are gathering and managing catalogues and data collections from multiple data and data product originators; the Blue-Cloud VRE is also considered as Blue Cloud data source, concerning publishable data products as resulting from the demonstrator Virtual Labs.

- **Marine-ID service**, operated by IFREMER, for registration and authentication of users to the Blue Cloud Data Discovery and Access service. Users only have to register once to receive their login details.

- **Data cache**, operated by EUDAT, for temporary storage of data packages, consisting of data sets, retrieved from the Blue Cloud data sources, plus associated metadata, as retrieved from the Blue Cloud data brokerage, and following the instructions as received from the Blue Cloud data brokerage. External users can download these data sets, after receiving information from the Blue Cloud data brokerage, while the VRE can also be triggered to retrieve data packages for ingestion into the catalogue and data pool of the Blue Cloud VRE.

- **Blue Cloud Data brokerage service**, operated by MARIS. This service performs the master role in the Blue Cloud Data Discovery and Access service, interacting with the other modules. Regularly, it retrieves the latest Blue Cloud level 1 metadata catalogue from the Blue Cloud metadata brokerage, and ingests this into the discovery interface, whereby users can query the catalogue at level 1. The common level 1 metadata catalogue includes sufficient metadata for each blue data source to allow the first level queries at collection level with a few selection criteria and this way to identify which of the blue data infrastructures holds interesting data sets. The Blue Cloud level 1 metadata catalogue should also contain sufficient additional metadata to allow more specific searching at level 2 for those blue data infrastructures that only have data collections and other data products, but no service at granule level. While for other blue data infrastructures, supporting deeper searching at level 2 – granule level -, customised search profiles have been formulated, which allow the data broker to interact with the provided web services and APIs of the blue data infrastructures. The Blue Cloud Data brokerage service also contains a shopping mechanism with basket and ledger, by which users (external users and VRE) and blue data infrastructures can be informed about shopping transactions and their status in time. It interacts with the Blue Cloud Data Cache to give it precise instructions about retrieving data sets from the blue data infrastructures and to insert these for temporary storage, and to bundle these as downloadable data packages for each shopping order. It interacts with the Marine-ID service as users need to login to submit shopping baskets and to have access to the transaction ledger. It interacts with registered users and VRE to inform and instruct them about data packages that are ready for downloading by users or retrieval for ingestion by the VRE. Finally, it also interacts with the Blue Cloud Data Cache to receive information about the actual downloading by users and retrieval for ingestion by the VRE in order to update the ledger.

- **User interfaces**, operated by MARIS, to interact with users for discovery and shopping transactions at level 2, and to provide access to the transaction ledger for users and blue data infrastructures.

3.7.5.3. Blue-Cloud Solutions applied to support JCORE Services

JCORE		Blue-Cloud Solution
Integration Layer	Catalogue of Services	
Community	JWP - JCORE Web Portals	Gateway - Web GUI
	VRE – Virtual Research Environment	Gateway - Virtual Research Environment Core - VRE Mgmt.
	TTC – Thematic & Technical Center	Gateway - Virtual Labs, VRE Core - VRE Mgmt.
	RPS – Resources Providers	Resource Providers - Cloud, Services, Information Systems
	TNA – Transnational Access	Community Application - Gateway, Collaborative Resource Providers - Cloud, Services, Information Systems
	JCA – JCore API	Not available, but is possible to develop ad hoc software running in Docker.
	SCN – SoCial Networking	Collaborative - Social Networking
Process	HRV – HarVesting	Data Analytics - Data Miner Data Management Space *Blue Cloud Data Discovery and Access Service*
	HYD – HYDra	Not available, but is possible to develop ad hoc software running in Docker.
	EVT - EvenT Engine	Not available, but is possible to develop ad hoc software running in Docker.
Information	VAMS – Virtual Access Metric System	Not available, but is possible to develop ad hoc software running in Docker.
	DAA - DAta Analysis	Data Analytics
	DAP - DAta Process	Data Analytics

	CTR - CaTalog of Resources	*Blue Cloud Data Discovery and Access Service*
Governance & System Management	AAI – Access & Auth. Infrastructure	Core - IAM
	SYS – SYStem management	Core - Information System
	GOV - GOVernance	Core - IAM, Resource Management, VRE Management
Application Platform	DBaaS - DataBase as a Service	*Blue Cloud Data Discovery and Access Service*
	IDaaS - IDentification as a Service	Performed by Blue-Cloud/D4Science
	PaaS - Platform as a Service	Performed by Blue-Cloud/D4Science
	IaaS - Infrastructure as a Service	Performed by Blue-Cloud/D4Science

3.7.6. Evaluation

This document concludes with a comparison of the explored e-infrastructures. This assessment has been based on criteria for the feasibility of the JERICO-CORE project development on the evaluated systems. Summarising these criteria topics briefly, they can be separated into two groups. In the first group, there are all those technical and functional viability criteria that give technical certainty about the successful development of JERICO-CORE. In the second group, there are criteria that evaluate the development difficulty and the technical and functional risks associated, which ultimately express a degree of uncertainty about the project development.

The next evaluation criteria is explained, detailed and scored on the evaluated platforms, always emphasising that they have been assessed from the perspective of the main purpose of the project: the integrated development of JERICO-CORE on already established research e-infrastructures.

Due to the technical nature of this document, criteria topics like economic and legal viability are not considered.

3.7.6.1 Evaluation Criteria

The next evaluation criteria topics and their respective aspects are evaluated using a uniform scale of 5 numbers from 0 to 4, and a code of colours, where:

Numeric Code	**Meaning**	
	In a Success Criteria	*In a Failure Criteria*
0	Very low	Very high
1	Low	High
2	Moderate	Moderate
3	High	Low
4	Very high	Very low

Color Code	**Meaning**
Black	Unacceptable risk
Red	Very high risk
Orange	High risk
Yellow	Risk moderated
Green	Desired situation

- **Technical viability**: This criteria refers to the technical capacity to build the system. Do you have the knowledge, skills and tools necessary to create the system? Is it possible to build it on other existing systems? Are there technical limitations that can prevent its development?

  The aspects considered in the Technical Viability criteria, are:

- **Availability of Tools and Technologies**: Evaluates the necessary tools and technologies to build the system and ensures they are available and suitable for the project. It considers aspects such as compatibility, stability and security of tools and technologies.

Scoring scale:

Score	Description
0	Very low availability: The necessary tools and technologies for the system are scarce or inadequate. The selection of tools is limited and may negatively impact the performance and functionality of the system.
1	Low availability: There are some available tools and technologies, but their variety and capability may be limited. Alternative or customised solutions may be required to meet specific system requirements.
2	Moderate availability: There is a good variety of tools and technologies available to support the system. Industry-standard solutions can be used and customised as needed.
3	High availability: There is a wide range of tools and technologies available that align with the system's requirements. Industry-standard solutions can be easily implemented, and additional options are available for advanced customizations.
4	Very high availability: There is a wealth of cutting-edge tools and technologies available to support the system. State-of-the-art solutions are accessible, and industry best practices are employed.

- **Technical Team Competencies**: Verifies that the team responsible for developing the system has the technical skills necessary to carry out the project. Make sure they have previous experience in similar projects and with the appropriate technical knowledge to handle the necessary tools and technologies.

Scoring scale:

Score	Description
0	Very low competencies: The technical team lacks the necessary skills and knowledge to develop, implement, and maintain the cloud system. Substantial training and a significant increase in technical competencies are required.
1	Low competencies: The technical team possesses some skills and knowledge, but they may be insufficient to address all technical aspects of the cloud system. Additional training and possibly the addition of new team members with relevant expertise are needed.
2	Moderate competencies: The technical team has the basic skills and knowledge required to manage and maintain the cloud system. Most technical tasks can be accomplished, but challenges may arise in specialised areas.
3	High competencies: The technical team has strong skills and knowledge in developing, implementing, and maintaining cloud systems. They can effectively address most technical challenges and are trained in the latest technologies and practices.
4	Very high competencies: The technical team is highly competent and has extensive experience in implementing and managing cloud systems. They can confidently address any technical challenge and stay up to date with the latest trends and advancements in the field.

- **Integration Capacity**: When building the system on other existing systems, it is important that you verify the system integration capacity you are building with existing systems. Should be ensured that there are no conflicts or incompatibilities that may affect the performance or stability of the system.

Scoring scale:

Score	Description
0	There is no possibility of integration. Systems are too much different or incompatible technically speaking. The e-infrastructures with this score in Integration Capacity, are not discarded due they have won high scores in Operational and Strategic viability criterias.
1	It is only possible to perform a partial integration, leaving out more than 50% of technical and/or functional requirements, which should be solved using other platforms, with a high uncertainty of how those platforms will be combined as a whole solution.
2	It is possible to achieve more than 80% of the technical and functional requirements, even combining different e-infrastructures. The rest of the technical and functional requirements should be possible to be developed inside the selected platforms.
3	It is possible to achieve more than 80% of the technical requirements using only one e-infrastructure. The rest of the technical requirements should be possible to be developed inside the selected platform.
4	It is possible to perform a complete or almost (>= 90%) total integration of all the technical requirements in one e-infrastructure with a minimum extra effort to develop some technical requirements inside this selected platform.

- **Risks and Technical Limitations**: Identifies possible risks and technical limitations that may affect the development of the project, such as the lack of technical resources, time limitations, complexity of the system, and the most important one: reaction capacity against changes in requirements, under the point of view of technical viability: Tools and technology should be effective and flexible enough. It is important to have a plan to face these risks and limitations.

Scoring scale:

Score	Description
0	Very high risk and limitations: The system faces multiple significant technical risks and limitations that can adversely affect its operation. There are substantial hurdles to overcome to ensure technical viability. The e-infrastructures with this score in risks and Technical Limitations, are not discarded due they have won high scores in Operational and Strategic viability criterias. However, this kind of e-infrastructures usually leads to a very radical change in the way in which users operate in order to "adapt" the use of the platform to the processes to which it is destined. This topic should be discussed and evaluated with representatives of different groups of users, and stakeholders.
1	High risk and limitations: The system faces some risks and limitations that may impact its functionality or performance. Mitigation strategies and additional measures may be necessary to address these challenges effectively.
2	Moderate risk and limitations: The system has moderate technical risks and limitations that can be managed and mitigated with proper planning and strategies. Some trade-offs may be required to ensure technical viability.
3	Low risk and limitations: The system has minimal technical risks and limitations that are well-understood and can be effectively managed. Mitigation plans are in place to address any potential issues.
4	Very low risk and limitations: The system has negligible technical risks and limitations, providing a solid foundation for technical viability. There are no significant obstacles that would impact the system's operation.

- **Scalability and Maintainability**: It realises that the system you are building is scalable and maintainable in the future. Verify that the system can grow and adapt to new needs and changes in the future, and that it can be maintained efficiently and sustainably.

Scoring scale:

Score	Description
0	Very low scalability and maintainability: The system has limited or no ability to scale effectively. It cannot handle increased demands or accommodate growth without significant limitations or performance degradation. Scaling requires extensive manual intervention or system redesign. The system is challenging to maintain and requires extensive effort to make modifications or updates. The components are poorly organised, and documentation is lacking. Tracking and troubleshooting issues is time-consuming and difficult. The e-infrastructures with this score in scalability and maintainability, are not discarded due they are based in robust long-term hardware and software technologies, and they have won high scores in Operational and Strategic viability criterias. The transactional bank systems are good examples of this kind of (legacy) platform.
1	Low scalability and maintainability: The system has limited scalability capabilities. It can handle some increases in demands or moderate growth, but with noticeable constraints or performance degradation. Scaling may require manual intervention or adjustments to accommodate changes. The system requires significant effort to maintain and make changes. The components may be complex or poorly structured, and documentation may be insufficient. Tracking and troubleshooting issues can be challenging, but with some additional effort, modifications can be made.
2	Moderate scalability and maintainability: The system demonstrates reasonable scalability capabilities. It can handle gradual increases in demands or moderate growth with minimal performance impact. Scaling can be achieved through manageable adjustments or configurations. The system has a reasonable level of maintainability. The components are structured and organised, and documentation exists to guide maintenance tasks. Tracking and troubleshooting issues are manageable, and modifications can be made with relative ease.
3	High scalability and maintainability: The system exhibits strong scalability capabilities. It can handle substantial increases in demands or significant growth with minimal performance impact. Scaling is achieved through efficient and automated mechanisms, allowing the system to adapt seamlessly. The system is highly maintainable. The components are well-structured and follow best practices, and comprehensive documentation is available. Tracking and troubleshooting issues are straightforward, and modifications can be made efficiently.
4	Very high scalability and maintainability: The system possesses exceptional scalability capabilities. It can handle exponential increases in demands or rapid and extensive growth without significant performance impact. Scaling is automatic, dynamic, and effortless, allowing the system to effortlessly accommodate any changes in workload or user demands. The system has excellent maintainability. The components are robust, modular, and well-documented, facilitating ease of maintenance and updates. Tracking and troubleshooting issues are minimal, and modifications can be made effortlessly.

- **Operational viability**: This criteria refers to the ability to operate the system once it is built. Is system sustainability taken into account? Is the human resources necessary to operate and maintain the system? Has the scalability of the system been considered?

  The operational viability criteria refers to the ability of the proposed system to be effectively and efficiently managed and maintained in the long term. This criterion takes into account aspects such as the availability of technical support, the availability of resources and personnel, the system's scalability, and the ease of system upgrades and maintenance.

  A system that lacks operational viability can result in increased downtime, poor system performance, and higher maintenance costs, which can negatively impact the success of the project. Therefore, it is crucial to consider the operational viability of a system when evaluating its feasibility.

  In the case of the JERICO-CORE project, the operational viability criteria would involve assessing the ability of the proposed system to be managed and maintained by the supporting research e-infrastructures where it will be deployed. This requires a further and deeper evaluation of the availability of technical support and resources, as well as the scalability of the system to ensure that it can accommodate the needs of a growing user base. Additionally, it would involve assessing the ease of system upgrades and maintenance to minimise downtime and ensure efficient performance.

  The aspects considered in the Operational viability criteria, are:

- **Scalability**: The proposed system should be scalable, meaning that it can accommodate changes in user demand and usage patterns over time. This involves assessing the system's ability to handle increasing volumes of data, users, and processing requirements, and the potential for future upgrades and enhancements.

Scoring scale:

Score	Description
0	Very low scalability: The system lacks the ability to scale effectively. It cannot handle increased demands or accommodate growth without significant limitations or performance degradation. Scaling requires extensive manual intervention or system redesign.
1	Low scalability: The system has limited scalability capabilities. It can handle some increases in demands or moderate growth, but with noticeable constraints or performance degradation. Scaling may require manual intervention or adjustments to accommodate changes.
2	Moderate scalability: The system demonstrates reasonable scalability capabilities. It can handle gradual increases in demands or moderate growth with minimal performance impact. Scaling can be achieved through manageable adjustments or configurations.
3	High scalability: The system exhibits strong scalability capabilities. It can handle substantial increases in demands or significant growth with minimal performance impact. Scaling is achieved through efficient and automated mechanisms, allowing the system to adapt seamlessly.
4	Very high scalability: The system possesses exceptional scalability capabilities. It can

Score	Description
	handle exponential increases in demands or rapid and extensive growth without significant performance impact. Scaling is automatic, dynamic, and effortless, allowing the system to effortlessly accommodate any changes in workload or user demands.

- **Reliability and availability**: The proposed system should be reliable and available, meaning that it can operate continuously and without interruption. This involves assessing the system's resilience to hardware and software failures, its ability to recover from errors and downtime, and the availability of technical support and maintenance services.

Scoring scale:

Score	Description
0	Low reliability and availability: The system has frequent failures and prolonged downtime, significantly affecting normal operation. Fault recovery is slow and unreliable.
1	Moderate reliability and availability, but fault recovery is slow and unreliable.
2	Moderate reliability and availability: The system has occasional failures and brief periods of downtime. Fault recovery is generally effective, but there may be some minor interruptions in operation.
3	High reliability and availability: The system rarely experiences failures and has minimal downtime. Fault recovery is fast and effective, and normal system operation is maintained in most cases
4	Very high reliability and availability: The system is highly reliable and practically experiences no failures or downtime. Fault recovery is nearly instantaneous, and normal operation is virtually uninterrupted.

- **Security and privacy**: The proposed system should be secure and protect user privacy, meaning that it can prevent unauthorised access, protect against data breaches, and ensure user confidentiality. This involves assessing the system's security measures, including encryption, access controls, and authentication protocols, and ensuring compliance with relevant data protection regulations.

Scoring scale:

Score	Description
0	Very low security and privacy: The system has severe vulnerabilities and lacks essential security measures. There is a high risk of unauthorised access, data breaches, and privacy violations. Sensitive information is inadequately protected.
1	Low security and privacy: The system has significant vulnerabilities and limited security measures. There is a notable risk of unauthorised access or data breaches, and privacy may be compromised to some extent. Additional security measures are needed.
2	Moderate security and privacy: The system has basic security measures in place but may still have some vulnerabilities. There is a moderate risk of unauthorised access or data breaches, although efforts are made to protect sensitive information.
3	High security and privacy: The system has robust security measures implemented to protect against common threats. There is a low risk of unauthorised access or data breaches. Sensitive information is properly encrypted and safeguarded.
4	Very high security and privacy: The system has advanced security mechanisms and follows industry best practices. It undergoes regular security audits and assessments. There is an extremely low risk of unauthorised access, data breaches, or privacy violations. Strong encryption and advanced security protocols are employed to protect sensitive information.

- **Usability**: The proposed system should be user-friendly and easy to use, meaning that it can be easily adopted and used by its intended users. This involves assessing the system's interface design, user experience, and training and support resources.

Scoring scale:

Score	Description
0	Very low usability: The system is highly complex and difficult to navigate or understand. It requires extensive training and expertise to use effectively. Users encounter frequent errors and struggle to accomplish tasks. The interface is confusing and unintuitive.
1	Low usability: The system has some usability issues and may be challenging for users to navigate or understand. It requires some training or assistance to use effectively. Users may encounter occasional errors or difficulties in completing tasks. The interface could benefit from improvements to enhance user experience.
2	Moderate usability: The system is generally usable, but there are areas for improvement. Users can navigate and understand the system with moderate ease. It may require minimal training or guidance to use effectively. Users can accomplish most tasks, although they may occasionally face minor challenges. The interface is reasonably intuitive, but some enhancements could be implemented.
3	High usability: The system is highly usable and user-friendly. Users can navigate and understand the system easily without significant training or assistance. Tasks can be completed efficiently with minimal errors or obstacles. The interface is intuitive and well-designed, providing a smooth user experience.
4	Very high usability: The system is exceptionally user-friendly and intuitive. Users can effortlessly navigate and understand the system from the start, requiring no training or assistance. Tasks can be completed quickly and efficiently without errors or difficulties. The interface is highly intuitive, visually appealing, and optimised for a seamless user experience.

● ***Strategic viability***: This criteria refers to the capacity of the project to align with the strategic objectives and the needs of the organisation. Is the project aligned with the vision and strategy of the organisation?

The strategic viability criteria evaluates whether the proposed system aligns with the overall strategic goals and objectives of the organisation or community it serves. This criteria takes into account factors such as the relevance of the system to the organisation's mission, the potential impact of the system on the organisation's operations, and the level of support for the system from key stakeholders: It is closely related to how widely it covers functional requirements.

In the case of the JERICO-CORE project, the strategic viability criteria would involve evaluating the alignment of the proposed system with the overall strategic goals of the research community it serves. This would require assessing the relevance of the system to the community's research objectives, as well as the potential impact of the system on the community's research operations.

Additionally, it would involve evaluating the level of support for the system from key stakeholders of partner organisations. If the proposed system does not align with the strategic goals of the community or lacks support from key stakeholders, it may not be feasible to proceed with its development.

Overall, the strategic viability criteria ensures that the proposed system is not only technically and operationally viable but also aligns with the long-term strategic goals and objectives of the organisation or community it serves.

The strategic viability criteria is broken down into next key aspects:

- **Relevance to mission**: The proposed system should align with the overall mission and strategic goals of the organisation or community it serves. This involves evaluating the extent to which the system supports the core activities and objectives of the organisation or community.

Scoring scale:

Score	Description
0	Not relevant: The system does not align with the organisation's mission, goals, or objectives. It does not contribute to fulfilling the core purpose or strategic direction of the organisation.
1	Somewhat relevant: The system has some relevance to the organisation's mission, but its alignment is limited. It partially contributes to achieving the organisation's goals or objectives.
2	Moderately relevant: The system is moderately aligned with the organisation's mission. It contributes significantly to achieving the organisation's goals or objectives, although some aspects may require further improvement or alignment.
3	Highly relevant: The system is highly aligned with the organisation's mission. It plays a critical role in achieving the organisation's goals or objectives and directly supports its core purpose and strategic direction.
4	Extremely relevant: The system is fully aligned with the organisation's mission. It is integral to the organisation's success and is essential for accomplishing its goals or objectives.

- **Impact on operations**: The proposed system should have a positive impact on the operations of the organisation or community. This involves evaluating the potential benefits and drawbacks of the system, including how it will affect research productivity, collaboration, and innovation.

Scoring scale:

Score	Description
0	Transformational impact: The system has a transformative impact on the organisation's operations, leading to a complete overhaul of processes or workflows. It fundamentally changes how the organisation operates and requires significant resources and effort.
1	Significant impact: The system has a significant impact on the organisation's operations, requiring substantial changes to processes or workflows. Adapting to the system may require a considerable effort.
2	Moderate impact: The system has a noticeable impact on the organisation's operations, resulting in substantial changes to processes or workflows. However, these changes can be managed with moderate effort.
3	Minor impact: The system has a slight impact on the organisation's operations, requiring some adjustments or modifications to existing processes or workflows.
4	Negligible impact: The system has minimal impact on the organisation's operations. It does not significantly affect existing processes or workflows.

- **Support from Stakeholders**: The proposed system should have the support of key stakeholders, including funding agencies, industry partners, and other research organisations. This involves assessing the level of interest and investment from these stakeholders and their willingness to collaborate and support the development of the system.

Scoring scale:

Score	Description
0	Lack of support: The system lacks support from key stakeholders. There is little to no enthusiasm or endorsement for its implementation or usage.
1	Limited support: The system has limited support from some stakeholders, but overall endorsement is lacking. There may be scepticism or resistance to its implementation.
2	Moderate support: The system enjoys moderate support from stakeholders. While not unanimously endorsed, there is a reasonable level of enthusiasm and willingness to adopt and utilise the system.
3	Strong support: The system receives strong support from most stakeholders. There is a high level of endorsement and commitment to its implementation and usage.
4	Overwhelming support: The system has overwhelming support from all key stakeholders. There is unanimous enthusiasm and strong commitment to its implementation and usage.

- **Sustainability**: The proposed system should be sustainable in the long term, meaning it should be able to continue to operate effectively and efficiently beyond its initial development and implementation. This involves assessing the system's scalability, the availability of resources and technical support, and the potential for future upgrades and maintenance.

Scoring scale:

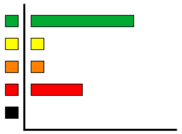
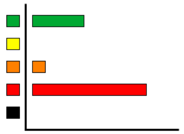
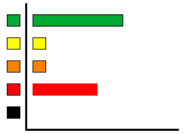
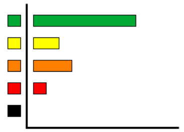
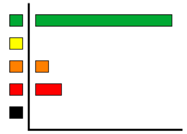
Score	Description
0	Not sustainable: The system is not sustainable in the long term. It lacks the necessary resources, capabilities, or plans to ensure its continued operation and viability.
1	Partially sustainable: The system has some sustainability measures in place but may face challenges in certain areas. It requires further efforts to enhance its long-term sustainability.
2	Moderately sustainable: The system demonstrates reasonable sustainability. It has adequate resources, capabilities, and plans to ensure its continued operation and viability.
3	Highly sustainable: The system is highly sustainable. It has robust measures in place to ensure long-term operation and viability, including sufficient resources, capabilities, and proactive planning.
4	Fully sustainable: The system is fully sustainable in the long term. It has comprehensive measures and strategies to ensure continued operation and viability, with ample resources, capabilities, and proactive planning.

- **Return on investment**: The proposed system should provide a positive return on investment, either through increased research productivity, collaboration, innovation, or other tangible benefits. This involves evaluating the potential costs and benefits of the system, including its potential impact on research outcomes and the wider community.

Scoring scale:

Score	Description
0	Negligible return: The system offers minimal or no return on investment. The benefits derived from the system do not justify the resources and effort invested.
1	Low return: The system provides limited return on investment. The benefits achieved are modest and may not fully justify the resources and effort invested.
2	Moderate return: The system delivers a reasonable return on investment. The benefits outweigh the costs and efforts, resulting in a satisfactory level of return.
3	High return: The system yields a significant return on investment. The benefits greatly outweigh the costs and efforts, resulting in substantial returns.
4	Exceptional return: The system generates an exceptional return on investment. The benefits far surpass the costs and efforts, resulting in outstanding returns and significant value for the organisation.

3.7.6.2 Evaluation Scoring

Evaluation Scoring (1/2)	egi	EUDAT	eosc	D4SCIENCE INFRASTRUCTURE	Blue-Cloud
Total Score	**37**	**26**	**32**	**39**	**45**
Risks					
Technical viability	**14**	**7**	**10**	**15**	**19**
- Availability of Tools and Technologies	3	1	2	3	4
- Technical Team Competencies	4	2	3	4	4
- Integration Capacity	1	1	1	2	4
- Risks and Technical Limitations	2	1	1	3	4
- Scalability and Maintainability	4	2	3	3	3

Evaluation Scoring (2/2)	EGI	EUDAT	eosc	D4SCIENCE INFRASTRUCTURE	Blue-Cloud
Operational viability	**14**	**12**	**12**	**12**	**11**
- Scalability	4	3	3	3	3
- Reliability and Availability	4	4	4	3	2
- Security and Privacy	4	4	4	4	4
- Usability	1	1	1	2	2
Strategic viability	**10**	**7**	**10**	**12**	**15**
- Relevance to Mission	2	1	2	3	4
- Impact on Operations	0	0	0	0	0
- Support from Stakeholders	1	1	1	2	3
- Sustainability	4	4	4	4	4
- Return on Investment	3	1	3	3	4

4. CONCLUSIONS

JERICO-RI is highly relevant in the field of oceanographic platforms. It is a European initiative that aims to enhance the capacity for ocean observation and monitoring in the coastal domain.

JERICO-RI promotes collaboration among different institutions and European countries to develop an integrated infrastructure for coastal ocean observation. This involves the implementation of a network of observatories and oceanographic platforms distributed across various marine regions in Europe.

The relevance of JERICO-RI lies in its ability to provide high-quality oceanographic data. These data are crucial for understanding and predicting changes in marine ecosystems, oceanic climate, and natural phenomena such as currents, salinity, and water temperature.

The JERICO-RI infrastructure enables the systematic collection of long-term data and provides open access to the scientific community and other interested users. This fosters research and the development of sustainable solutions related to ocean management, fisheries, marine environmental protection, and the mitigation of natural disasters.

JERICO-DS is an essential part of the JERICO-RI initiative dedicated to improving the quality and access to coastal and oceanographic data through a network of observatories and advanced networked services. Its goal is to foster collaboration and data exchange among European institutions and countries, thereby enhancing the understanding and management of coastal and marine ecosystems.

However, as presented in this document, nothing of this is complete without the proper federated cloud-computing infrastructure, where terms like federated FAIR data, efficient data processing, collaboration and accessibility are crucial.

The major aim of this JERICO-CORE Technical Design Study is to establish the proper fundations of the JERICO-RI e-infrastructure roadmap, basing its development on the advanced and major existing European Initiatives, gaining knowledge, experience and a quick return on investment in terms of costs and time to release.

5. OUTREACH, DISSEMINATION AND COMMUNICATION ACTIVITIES

6. ANNEXES AND REFERENCES

6.1. APPENDIX A - JERICO-RI Community Requirements

A set of requirements was collected from the needs of JERICO-S3 project and among JERICO-RI partners, modellers, product developers and other experts in collaboration with all the work packages. The resulting pilot was built considering existing landscape of infrastructures. Firstly, we avoid creating another resource repository but instead, existing data and document repositories such as EMODnet, SeaDataNet and the Ocean Best Practice System (OBPS) were used as information providers. Moreover, we collaborated with other large Research Infrastructures to integrate to their systems. This was motivated by the short amount of time and resources available for the development of the pilot. Two MoU were signed: (1) with the European Plate Observing System (EPOS) to use their solution to visualise JERICO resources; (2) with Blue-Cloud (BC) to make use of their VRE to create services within a JERICO Virtual Lab. The JERICO-CORE pilot is operated as a WP11 VA to support users and to identify gaps and needs for the design of a more sustained infrastructure under JERIC-DS WP3.

The next list of requirements, gathered from
JERICO-DS_MS.13-WP3_e-JERICO_requirements_compiled

The JERICO-DS project is funded by the European Commission's H2020 Framework Programme under grant agreement No. 951799. Project coordinator: Ifremer, France.

Requirements distributed by JERICO-CORE Modules															
Req. Category	Requirement	JWP	CTR	HRV	DAA	DAP	HYD	JCA	VAMS	RPS	TA	TTC	SCN	GOV	EVT
1. Distributed JERICO-RI infrastructures, including Resources catalogue	1. Online catalogue providing information on JERICO resources with link to data access	●	●					●							
	2. Catalogue with detailed metadata, instrument and parameter specific, and including QA information	●	●	●										●	
	3. Enhanced sensor metadata		●	●	●	●								●	
	4. Pilot Super Sites (PSS) and Integrated Regional Regional Sites (IRS) information		●	●											
	5. Information on deployments		●	●											
	6. Catalogue of developed and relevant software		●	●											
	7. Catalogue of Best Practices		●	●											
	8. Catalogue of tools and documents both existing and those developed within JERICO		●	●											
	9. Ensure metadata catalogue linked to Copernicus Marine Service INSTAC and EMODnet		●	●											

The JERICO-DS project is funded by the European Commission's H2020 Framework Programme under grant agreement No. 951799. Project coordinator: Ifremer, France.

	10. Forum to ask questions and receive answers from experts	●											●		
	11. A federated system	●	●					●			●		●	●	
	12. Information on TNA structures	●	●								●				
	13. Set up thematic technical centres											●			
	14. Close the oxygen data gap	●	●		●	●	●								
	15. Ensure visibility of partners in the RI	●											●		
	16. Identify and ensure coastal data not yet in EU data aggregators are made available and interoperable		●	●										●	
	17. Easy to find data and information for regions i.e. North Sea, Baltic,....	●	●	●	●		●								
		JWP	CTR	HRV	DAA	DAP	HYD	JCA	VAMS	RPS	TA	TTC	SCN	GOV	EVT
2. Thematic Services	18. JERICO toolboxes	●			●	●	●	●							
	19. Identify partner services that can be provided to e-JERICO and JERICO to combine/package the services										●				
	20. Good search engine with AI support	●	●	●	●		●	●					●	●	
	21. Provide customised end user applications	●						●			●		●	●	

The JERICO-DS project is funded by the European Commission's H2020 Framework Programme under grant agreement No. 951799. Project coordinator: Ifremer, France.

	22. Provide tool for regional Sea Conventions to produce products and to be the information source feeding the OSPAR operational objectives	●	●	●	●	●	●	●							
	23. Create a JERICO event alarm system (tsunamis, Baltic inflows, HABs,....)		●		●	●	●								●
	24. JERICO climatology	●	●	●	●	●	●	●					●	●	●
	25. Tools to provide platform operation status (based on technical centre requirements).	●	●	●								●			●
	26. Tool to track the data flow i.e. what data goes where		●	●											
		JWP	CTR	HRV	DAA	DAP	HYD	JCA	VAMS	RPS	TA	TTC	SCN	GOV	EVT
3. Custom User and Machine-to-Machine Interfaces	27. Flexible, easy to use and attractive	●											●		
	28. Provide service for any data originator to add data to the system via registration page. Share data and become part of the JERICO community			●			●	●			●	●			
	29. Provide a "new data" alert with notification for registered users of new data available	●	●	●											●
	30. Federated JERICO ERDDAP	●	●	●			●								
	31. Promote ERDDAP within the JERICO community	●	●	●			●						●		

	32. Ensure interoperability with Copernicus Marine Service INSTAC and EMODnet	●	●	●			●								
	33. Ensure JERICO data are labelled as "JERICO" to raise the visibility		●	●										●	
	34. Stakeholder product development support	●	●	●	●	●	●	●					●		
		JWP	CR	HV	DA	DP	HY	JCA	VAMS	RPS	TA	TTC	SN	GV	EVT
4.Custom Virtual Lab	35. Quality control tools with common protocols		●	●	●	●								●	
	36. Tool for intercomparison between platforms and ensure data collected from different platforms are interoperable	●	●	●	●	●	●	●						●	
	37. Provide tool for comparison observations vs models	●	●	●	●	●	●	●							
	38. Unified resource for coastal data	●	●	●			●	●							
	39. AI driven data warehouse providing 1st level data analysis tools	●	●	●	●		●	●							
	40. Tool to produce gridded data on the fly	●	●		●	●	●	●							
	41. Provide a data processing tool	●	●			●	●	●							
	42. Connect ocean to land via "earth dataweb" i.e. combining ocean and atmospheric data		●	●	●	●	●	●							

	43. Framework to use data together with providing tools to make added value products	●	●		●	●	●						●		
		JWP	CTR	HRV	DAA	DAP	HYD	JCA	VAMS	RPS	TA	TTC	SCN	GOV	EVT
5. Others	44. Provide assistance for near-real time presentations of data	●	●		●	●	●	●							
	45. JERICO act as a test bed for new sensors	●	●	●	●	●	●	●							
	46. Be clear with what JERICO delivers regarding systems and data i.e. define JERICO platforms/data		●	●										●	
	47. FAIR		●	●	●	●	●	●						●	
	48. Seek synergies with other RIs	●		●			●						●		
	49. More focus on Biology		●	●											
	50. Set EU standards for biological data		●	●	●	●	●							●	
	51. Take lead for biological Best Practices		●		●								●	●	
	52. Identify JERICO users														
	53. Link research to operational oceanographic needs										●	●	●		
	54. Identify area covered i.e. "coastal"		●	●											
	55. Data and metadata reaching certain quality level to receive		●	●	●	●	●	●						●	

	a JERICO label as a quality stamp/assurance to users														
	56. Leverage on what's done elsewhere avoid duplication of efforts		●												
	57. Ensure standards, etc. set in other RIs go through the full value chain and ensure interoperability		●	●	●	●	●	●						●	
	58. Provide statistics to data providers on use of data								●	●					
	59. Not another data portal, avoid duplication with EU data aggregators		●												
	60. Assist in developing assessments for the regional sea conventions	●	●		●	●									
	61. JERICO adjust to the needs of regional assessments														
	62. JERICO needs to identify and show what it can deliver		●	●	●	●	●							●	
	63. Promote JERICO within the regional sea conventions								●				●	●	
	64. Act as a new technology forum	●	●		●	●	●	●		●	●	●	●		●

6.2. *Glossary of terms*

JERICO-CORE Service Architecture

- Community Integration Layer
  - JWP — JCore Web Portals
  - VRE — Virtual Research Environment
  - TTC — Thematic & Technical Centers
  - RPS — Resource Providers
  - TA — Transnational Access
  - JCA — JCore API
- Process Integration Layer
  - HRV — HaRVesting
  - HYD — HYDra
  - EVT — EVenT Engine
- Information Integration Layer
  - VAMS — Virtual Access Metric System
  - DAA — DAta Analysis
  - DAP — DAta Process
  - CTR — CaTalog of Resource
- Governance & System Management Layer
  - AAI — Access & Authorization Infrastructure
  - SYS — SYStem Management
  - GOV — GOVernance

6.3. References

[1] ESFRI Roadmap 2021 - Strategy Report on Research Infrastructures.
https://roadmap2021.esfri.eu/media/1295/esfri-roadmap-2021.pdf

[2] ESFRI Roadmap 2021 - Strategy Report on Research Infrastructures - Public Guide.
https://www.esfri.eu/sites/default/files/ESFRI_Roadmap2021_Public_Guide_Public.pdf

[3] Towards Interoperable Research Infrastructures for Environmental and Earth Sciences, A Reference Model Guided Approach for Common Challenges

https://link.springer.com/book/10.1007/978-3-030-52829-4

Wilkinson, M.D., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Bonino da Silva Santos, L., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18 (2016).

[4] JERICO-S3 D11.1 First report on VA JERICO Resources access
https://drive.google.com/file/d/1b1fFv4dfpTakUFKVen52RKXiDpREMRuZ/view

[5] Reference framework for assessing the scientific and socio-economic impact of research infrastructures" OECD
https://doi.org/10.1787/3ffee43b-en

[6] D. Chadwick, "Federated Identity Management," in Foundations of Security Analysis and Design V SE - 3, vol. 5705, A. Aldini, G. Barthe, and R. Gorrieri, Eds. Springer Berlin Heidelberg, 2009, pp. 96–120.

[7] ISO/IEC 17788:2014: Information Technology – Cloud Computing — Overview and Vocabulary.
https://www.iso.org/standard/60544.html

[8] ISO/IEC 17789:2014: Information technology — Cloud computing — Reference architecture.
https://www.iso.org/standard/60545.html

[9] Cloud Computing Governance Framework - The Open Group.
http://www.opengroup.org/cloud/gov_snapshot/p1.htm

[10] The EGI applications on Demand service
https://doi.org/10.1016/j.future.2019.03.009

[11] Earth Observation Open Science and Innovation, ISSI Scientific Report Series 15
https://doi.org/10.1007/978-3-319-65633-5_3

[12] e-Infrastructure Reflection Group Support Programme 7 (e-IRGSP7)
https://e-irg.eu/e-irgsp7/

[13] EUDAT Collaborative Data Infrastructure / Projects
https://eudat.eu/projects

[14] EUDAT Collaborative Data Infrastructure / Service catalogue
https://eudat.eu/catalogue

[15] European Commission (2015) Open Science at the Competitiveness Council.
http://ec.europa.eu/digital-agenda/en/news/open-science-competitiveness-council-28-29-may-2015

[16] European Commission (2016) First report of High Level Expert Group on the EOSC.
https://ec.europa.eu/digital-single-market/en/news/first-report-high-level-expert-group-european-open-science-cloud

[17] Tinman - Solutions for a sustainable EOSC
https://www.eoscsecretariat.eu/system/files/solutions_for_a_sustainable_eosc_-_tinman_draft_02dec19.pdf

[18] FAIR Lady report
https://op.europa.eu/en/publication-detail/-/publication/581d82a4-2ed6-11eb-b27b-01aa75ed71a1

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

[20] EOSC Federating Community Position Paper
https://eosc-hub.eu/sites/default/files/EOSC%20Federating%20Core%20Community%20Position%20Paper%20v1.1.pdf

[21] Achieving Open Science in the European Open Science Cloud: Setting out OpenAIRE's vision and contribution to EOSC
https://doi.org/10.5281/zenodo.3475076

[22] EOSC Interoperability Framework
https://www.afs.enea.it/project/madia/Documenti/web/docs/EOSC-Interoperability-Framework-KI0221055ENN.pdf

[23] EOSC Architecture and Interoperability Framework
https://eosc-portal.eu/sites/default/files/EOSC%20Future-WP3-EOSC%20Architecture%20and%20Interoperability%20Framework-2021-12-22%5B17%5D%5B6%5D-2.pdf

[24] EOSCpilot - WP5 - D5.4 Final EOSC Service Architecture
https://eoscpilot.eu/sites/default/files/eoscpilot_d5.4-v1.1.pdf

[25] EOSCpilot - WP5 - D5.1 Initial EOSC Service Architecture
https://eoscpilot.eu/sites/default/files/eoscpilot-d5.1.pdf

[26] European Commission (2018) Prompting an EOSC in practice: Final report and recommendations of the Commission 2nd High Level Expert Group on the European Open Science Cloud (EOSC). EU Publications, Directorate General for Research and Innovation. DOI: 10.2777/112658

[27] DILIGENT -A testbed digital library infrastructure on grid enabled technology
Programme(s) FP6-IST - Information Society Technologies: thematic priority under the specific programme "Integrating and strengthening the European research area" (2002-2006).
https://cordis.europa.eu/project/rcn/71926/en

[28] D4Science - DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science (2008-2009 - EU Grant agreement No. 212488)
https://cordis.europa.eu/project/id/212488

[29] D4Science-II - Data Infrastructure Ecosystem for Science (2009-2011 - EU Grant agreement No. 239019)
https://cordis.europa.eu/project/id/239019

[30] ENVRI - Common Operations of Environmental Research Infrastructures (2011- 2014 - EU Grant agreement No. 283465)
https://cordis.europa.eu/project/id/283465

[31] EUBrazilOpenBio - EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity (2011-2013 - EU Grant agreement No. 288754)
https://cordis.europa.eu/project/rcn/99564/en

[32] iMarine - Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources (2011-2014 - EU Grant agreement No. 283644)
https://cordis.europa.eu/project/rcn/99770/en

[33] BlueBRIDGE - Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth (2015-2018 - EU Grant agreement No. 823914)
https://cordis.europa.eu/project/rcn/198301/en

[34] EGI-Engage - Engaging the EGI Community towards an Open Science Commons (2015-2017 - EU Grant agreement No. 654142)
https://cordis.europa.eu/project/rcn/194937/en

[35] ENVRI PLUS - Environmental Research Infrastructures Providing Shared Solutions for Science and Society (2015-2019 - EU Grant agreement No. 654182)
https://cordis.europa.eu/project/rcn/194947/en

[36] PARTHENOS - Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies (2015-2019 - EU Grant agreement No. 654119)
https://cordis.europa.eu/project/rcn/194932/en

[37] SoBigData - SoBigData Research Infrastructure (2015-2019 - EU Grant agreement No. 654024)
https://cordis.europa.eu/project/rcn/197882/en

[38] AGINFRA PLUS - Accelerating user-driven e-infrastructure innovation in Food Agriculture (2017-2019 - EU Grant agreement No. 731001)
https://cordis.europa.eu/project/rcn/206341/en

[39] PerformFISH - Consumer driven Production: Integrating Innovative Approaches for Competitive and Sustainable Performance across the Mediterranean Aquaculture Value Chain (2017-2022 - EU Grant agreement No. 727610)
https://cordis.europa.eu/project/rcn/210496/en

[40] ARIADNE PLUS - Advanced Research Infrastructure for Archaeological Data Networking in Europe - plus (2019-2022 - EU Grant agreement No. 731001)
https://cordis.europa.eu/project/id/823914

[41] EOSC Pillar - Coordination and Harmonisation of National Initiatives, Infrastructures and Data services in Central and Western Europe
https://cordis.europa.eu/project/id/857650

[42] DESIRA - Digitisation: Economic and Social Impacts in Rural Areas (2019-2023 - EU Grant agreement No. 818194)
https://cordis.europa.eu/project/id/818194

[43] RISIS2 - European Research Infrastructure for Science, technology and Innovation policy Studies 2 (2019-2022 - EU Grant agreement No. 824091)
https://cordis.europa.eu/project/id/824091

[44] Blue-Cloud - Piloting innovative services for Marine Research & the Blue Economy (2019-2022 - EU Grant agreement No. 862409)
https://cordis.europa.eu/project/id/862409

[45] SoBigData++ - European Integrated Infrastructure for Social Mining and Big Data Analytics (2020-2023 - EU Grant agreement No. 871042)
https://cordis.europa.eu/project/id/871042

[46] D4Science: an e-Infrastructure for Supporting Virtual Research Environments.
Leonardo Candela, Donatella Castelli, and Pasquale Pagano.
In Maristella Agosti, Floriana Esposito, and Costantino Thanos, editors, Post-proceedings of the 5th Italian Research Conference on Digital Libraries - IRCDL 2009, pages 166–169. DELOS: an Association for Digital Libraries, 2009.

[47] Realising Virtual Research Environments by Hybrid Data Infrastructures: the D4Science Experience
http://dx.doi.org/10.13140/2.1.5035.2327
Leonardo Candela, Donatella Castelli, and Pasquale Pagano
Conference: International Symposium on Grids and Clouds (ISGC) 2014, Proceedings of Science PoS(ISGC2014)022At: Academia Sinica, Taipei, Taiwan

[48] gLite Lightweight Middleware for Grid Computing.
http://glite.web.cern.ch/glite/, https://glite.web.cern.ch/glite/introduction

[49] EGEE: Enabling Grids for E-sciencE.
http://www.eu-egee.org, https://eu-egee.org/fileadmin/documents/publications/EGEEIII_Publishable_summary.pdf

[50] gCube v1.0: A Software System for Hybrid Data Infrastructures. Technical Report 2008-TR-035,
https://www.researchgate.net/publication/267539466_gCube_v10_A_Software_System_for_Hybrid_Data_Infrastructures
Leonardo Candela, Donatella Castelli, and Pasquale Pagano
Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", CNR, 2008.

[51] gCube: A Service-Oriented Application Framework on the Grid.
L. Candela, D. Castelli, and P. Pagano.
ERCIM News, (72):48–49, January 2008.
https://www.researchgate.net/publication/200462065_gCube_A_Service-Oriented_Application_Framework_on_the_Grid

[52] D2.7 Blue Cloud Architecture (Release 2) - WP2, developing the Blue Cloud Discovery and access service and overall Blue Cloud architecture.
https://zenodo.org/record/6334925#.ZBglmtLMJD8

Dick M.A. Schaap (MARIS), Peter Thijsse (MARIS),
Pasquale Pagano (CNR-ISTI), Massimiliano Assante (CNR-ISTI), Leonardo Candela (CNR-ISTI),
Enrico Boldrini (CNR-IIA),
Merret Buurman (DKRZ),
Mattia d'Antonio (CINECA),
Chris Ariyo (CSC),
Gilbert Maudire (IFREMER), Cecile Nys (IFREMER)
Blue-Cloud: Piloting innovative services for Marine Research & the Blue Economy
Horizon programme call BG-07-2019-2020, topic: [A] 2019 - Blue Cloud services, Grant Agreement n. 862409.

https://cordis.europa.eu/project/id/862409