Project Title:	ECOPOTENTIAL: IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Project number:	641762
Project Acronym:	ECOPOTENTIAL
Proposal full title:	IMPROVING FUTURE ECOSYSTEM BENEFITS THROUGH EARTH OBSERVATIONS
Type:	Research and innovation actions
Work program topics addressed:	SC5-16-2014: "Making Earth Observation and Monitoring Data usable for ecosystem modelling and services"

Deliverable No: D10.1

Design of the ECOPOTENTIAL Virtual Laboratory

Version:	V1.0
Main Authors:	Stefano Nativi, Paolo Mazzetti, Mattia Santoro

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 641762

Outline

Executive Summary

The ECOPOTENTIAL project aims to build a unified framework for ecosystem studies and management of protected areas. To achieve such objective, the open and interoperable access to data and knowledge is assured by a GEO Ecosystem Virtual Laboratory Platform, fully integrated in GEOSS. The concept of GEO Ecosystem Virtual Laboratory stems from the need of moving from open data to open science as a new vision of participatory scientific research. Therefore, it aims not only to data sharing but more generally to provide support the ecosystem community-of-practice in research activities.

The architecture of the GEO Ecosystem Virtual Laboratory is based on a set of principles currently shared in the scientific research communities, with particular reference to the GEO initiative, including GEOSS Data Sharing Principles, GEOSS Data Management Principles and GEOSS Architecture Principles. Moreover, since ECOPOTENTIAL participates in the Horizon 2020 pilot action on open access to research data, the activities of the ECOPOTENTIAL Consortium for the definition of the ECOPOTENTIAL Data Management Plan are a fundamental input for the architecture of the ECOPOTENTIAL Virtual Laboratory.

The design of the ECOPOTENTIAL Virtual Laboratory puts its basis on past experiences in building System of Systems through a brokering approach. In brokered architectures, dedicated components provide mediation and harmonization of interfaces and data models avoiding the need of changes in the data provider systems. The mature data brokering approach will be complemented with innovative semantic technologies – including concept-based queries and annotations – and support of discovery and invocation of workflows implementing storylines on multiple protected areas contributing to enable the open science vision in ecosystem science.

The development focuses on loosely-coupled integration of mature technologies and tools, based on open tools or components provided or under control of the ECOPOTENTIAL Consortium members. The integration of new tools in the Virtual Laboratory is based on full server-side APIs, while applications development is facilitated through simple client-side APIs based on widespread Web technologies (HTML5, Javascript and CSS).

For greater flexibility, ECOPOTENTIAL adopts an agile methodology allowing rapid development in response to new requirements. It will have four main yearly iterations with fixed objectives for demonstration in reviews and events.

1 INTRODUCTION

This document describes the system architecture of the ECOPOTENTIAL Virtual Laboratory: a service-based platform for a virtual (i.e. online distributed) and open (i.e. accessible) laboratory to study ecosystems and contribute to GEO/GEOSS for facilitating access to Open Data.

The present version bases on: (a) the general context of cyberinfrastructures and Virtual Research Environments (VRE) in multidisciplinary science, (b) the requirements from the *H2020-SC5-2014-two-stage* call on the *SC5-16-2014* topic of the Societal Challenge 5, and (c) specific requirements and constraints collected during the preparation of the ECOPOTENTIAL Data Management Plan (DMP). Since, at the moment of preparation of the present deliverable (May 2016), the DMP is still to be delivered in its final form, due to the missing provision of relevant information from resource providers, the system architecture will be revised during the course of the project, with yearly releases of deliverable updates.

This document is the System Definition Document as described in the IEEE Guide to the Software Engineering Body of Knowledge, aiming at listing "*the system requirements along with background information about the overall objectives for the system, its target environment, and a statement of the constraints, assumptions, and non-functional requirements*" [1]. Although the development phase will be carried out inside the Consortium, therefore without the need to establish any formal "*agreement between customers and contractors or suppliers*" which are the objective of System Requirements Specification and Software Requirement Specification, some related information is provided when considered needed or useful.

After this Introduction, a second section focuses on the objectives and rationale behind the project, clarifying the main relevant concepts for ECOPOTENTIAL, such as what Open Data, Knowledge and Science, and providing an operational definition of Virtual Laboratory.

A third section reports an analysis on actors, user requirements and system requirements.

The fourth section describes the ECOPOTENTIAL architectural principles, focusing specifically on the need of loosely coupled applications, and on the brokering approach which is at the core of the ECOPOTENTIAL Virtual Laboratory concept.

The fifth section describes the ECOPOTENTIAL system architecture according to the viewpoint modelling approach through the five views defined by the Reference Model for Object Distributed Processing from ISO (RM-ODP).

A sixth section introduces the agile development approach that is adopted by the ECOPOTENTIAL project, and the sixth and final section reports the deployment plan and achievements at project-month 12.

2 RATIONALE AND MAIN CONCEPTS

2.1 The ECOPOTENTIAL main objective

The ECOPOTENTIAL project is funded in the Horizon 2020 EU Framework Programme for Research and Innovation (H2020). In particular, it was proposed as a response to the *H2020-SC5-2014-two-stage* call on *SC5-16-2014* topic of the Societal Challenge 5 (Climate action, environment, resource efficiency and raw materials) for "Making Earth Observation and Monitoring Data usable for ecosystem modelling and services" [2].

The key statement of the call says that:

Proposals should focus on recovering existing data, supporting new measurements and observations, synthesis and interpretation of data for making all information and knowledge available to scientists, policy makers, citizens and other concerned stakeholders to provide a full picture of the state and temporal evolution of ecosystems in existing internationally recognised protected areas.

It identifies the goal of the project as *to provide a full picture of the state and temporal evolution of ecosystems* and the operational objective as *making all information and knowledge available to scientists, policy makers, citizens and other concerned stakeholders*. The present document describes the system architecture for reaching the operational objective.

2.2 The ECOPOTENTIAL context and conditions

The ECOPOTENTIAL context poses some significant conditions on how the operational objective must be fulfilled. In particular, the *H2020-SC5-2014-two-stage* call on *SC5-16-2014* topic, directly or indirectly, establishes a set of major constraints:

C1. Utilisation of GEOSS, Copernicus and ESA data where possible. (*All activities under Societal Challenge 'Climate action, environment, resource efficiency and raw materials' should as far as possible use data resulting from or made available through different initiatives of the European Commission. In particular, the utilisation of GEOSS (Global Earth Observation System of Systems) and Copernicus (the European Earth Observation Programme) data, products and information should be privileged. Likewise, in line with EU cooperation with the European Space Agency (ESA), activities should use ESA Earth Science data, as far as possible. The data, both from ESA missions or third party missions, are for the vast majority of cases available for free web download*. [3])

C2. Participation in the Pilot on Open Research Data in Horizon 2020. (*The projects funded under Societal Challenge 'Climate action, environment, resource efficiency and raw materials', call 'Growing a Low Carbon, Resource Efficient Economy with a Sustainable Supply of Raw Materials', of the Work Programme 2014-15, with the exception of topics SC5-11-2014/2015, SC5-12-2014/2015, and SC5-13/2014/2015, will participate in the Pilot on Open Research Data in Horizon 2020 in line with the Commission's Open Access to research data policy for facilitating access, re-use and preservation of research data.* [3])

C3. Compliance with GEOSS Data Sharing Principles [4]. (*Beneficiaries in projects participating in the Pilot on Open Research Data shall adhere to the GEOSS Data Sharing Principles.* [3])

C4. Registration in GEOSS. (*Beneficiaries in projects participating in the Pilot on Open Research Data shall […] undertake to register in GEOSS all geospatial data, metadata and information generated as foreground of the project.* [3])

C5. Services based on improved access to GEOSS. (*…new prototype products and ecosystem services, based on improved access to (notably via GEOSS)…* [2])

C6. Long-term preservation of products. (*…longterm storage of ecosystem Earth Observation data and information in existing protected areas...* [2])

The formal aspects from C2 (Participation in the Pilot on Open Research Data in Horizon 2020) add further conditions:

C7. Open and free access to foreground data and possibly tools. (*Regarding the digital research data generated in the action ('data'), the beneficiaries must: (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following: (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; (ii) other data, including* *associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1); (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).* [5])

C8. Definition of a Data Management Plan and compliance with it [6]. (*The use of a Data Management Plan is required for projects participating in the Open Research Data Pilot.* [3])

2.3 Geospatial information in ECOPOTENTIAL: role and issues

The study of ecosystems deeply bases on geospatial information that is "information concerning phenomena implicitly or explicitly associated with a location relative to the Earth" [7]. Geographic Information is represented and conveyed through (geo)spatial data that is "any data with a direct or indirect reference to a specific location or geographical area" [8].

The geoinformation world is characterized by great complexity with many actors involved including:

- *Data (and information) producers* who acquire observations (e.g. through sensors) or generate value-added information (e.g. through data processing);
- *Data providers* who distribute data, managing data centres, long-term preservation archives, Spatial Data Infrastructures, etc.
- *Overarching initiatives* that influence the geoinformation world, designing new solutions, building disciplinary or interdisciplinary systems of systems, managing high-level expert groups, etc.
- *Technology providers* who develop and distribute technological solutions for geospatial data management and sharing
- *Cloud providers* who manage complex infrastructures on behalf of other actors such as data providers or application developers
- *Application developers* who make use of data to build applications for end-users
- *End-users* who utilize data

In such a context, interoperability is clearly perceived as one of the main issues even limiting to technological aspects. Indeed, actions of actors have an impact in terms of technological choices (see Figure 1).

- *Data (and information) producers* are mostly focused on data and metadata models and formats. Multiple standards have been defined addressing issues which are specific for different disciplinary domains, such as HDF, netCDF and GRIB for EO data, ESRI Shapefile or OGC GML for feature type information. Proprietary formats are still widespread;
- *Data providers* are mainly focused on data sharing services. As for data models and formats, several standards have been designed and adopted in different disciplinary domains. For example, in the biodiversity context TDWG standards are widely adopted, in the meteo-ocean community THREDDS Data Server is a widespread technology. OGC standard services are commonly adopted in the GIS community. Light specifications like KML (now an OGC standard) or OpenSearch are also common. OAI-PMH is a standard for long-term preservation archives.
- *Overarching initiatives* influence technological aspects in several ways, in particular on data management (e.g. the Data Management Plan guidelines in H2020 programme), data harmonization (e.g. WMO information systems specifications) and data sharing, including policy (e.g. RDA).
- *Technology providers* contribute to the heterogeneity providing many different competing solutions for geospatial data sharing. While some of them have adoption of standards as an objective, others (often from big players) prefer to push their own proprietary solutions.

- *Cloud providers* affect technologies providing new data storage and processing capabilities requiring new solutions for integration with traditional systems.
- *Application developers* contribute to the heterogeneity of the geoinformation world because they provide geospatial applications adopting different technologies, from operating systems and related ecosystems (e.g. Linux, Microsoft, Apple, Google Android), to development platforms (e.g. Java, Python, Javascript) and libraries.

Figure 1 Technological heterogeneity in the geoinformation world

The *H2020 SC5-16-2014* call explicitly mentions this issue saying that "*there is a need to develop innovative solutions that will provide open and unrestricted access to <u>interoperable ecosystem Earth Observation data and information</u>*". This statement also highlight that innovative solutions are required, since the definition of a unique standard is not a viable solution. Indeed, the lack of a standard is more the consequence of the complexity of the geospatial world than the reason of it. Including many actor categories, many disciplines, and many stakeholders (public authorities, private companies, citizens, etc.) **the complexity of the geospatial world makes impossible to agree on a single (or a small set) of standards** and, later, impose and enforce their adoption.

2.4 Open Data in ECOPOTENTIAL

It is recognized that there is a lack of clarity about key terms in literature and public debates related to Open Data [9]. In particular, the ambiguity of widely-used terms like "open" and "free" has caused misunderstanding, mixing-up concepts like "free usage" and "free of charge", and consequently nourishing the *gratis* (i.e. for zero price) vs. *libre* (i.e. with little or no restriction) debate. The Open Definition, from the Open Knowledge non-profit network, "*makes precise the meaning of 'open' with respect to knowledge, promoting a robust commons in which anyone may participate, and interoperability is maximized.*" It bases on the assumption that knowledge "*is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness*". It is explicitly clarified that, in this definition, "free" matches the "*libre*" concept [10].

Concerning ECOPOTENTIAL, the call provides few hints limiting to state that "*there is a need to develop innovative solutions that will provide <u>open and unrestricted access</u> to interoperable ecosystem Earth*

Observation data and information" [2]. Although this definition helps to clarify the data typology (i.e. Earth Observation data and information), it actually reiterates the *gratis* vs. *libre* ambiguity concerning policy: it does not specify whether "open and unrestricted" should be meant as "with little or no restriction" (*libre*) or "for zero price" (*gratis*).

The main source of information about how Open Data must be considered in ECOPOTENTIAL is the Grant Agreement establishing the rules of participation to the Horizon 2020 pilot action on open access to research data. The article 29.3 on "Open access to research data" explicitly states that "*Regarding the digital research data generated in the action ('data'), the beneficiaries must: (a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following: (i) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; (ii) other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1); (b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves)*" [5]. It provides clear conditions for all data generated in ECOPOTENTIAL: they must be deposited in a repository and made accessible free of charge to any user, as soon as possible (if they are used in scientific publications) or within a deadline defined in the data Management Plan (if they are not used in scientific publications). Moreover, the tools and instruments necessary to validate the scientific publications must be described and, where possible, provided. This is in line with the concept of Open Science and in particular of science reproducibility.

To comply with these requirements, as part of the WP1 (Coordination and management) activities, ECOPOTENTIAL has started the definition of its Data Management Plan (DMP) following the "Guidelines for Data Management in Horizon 2020" [6].

2.5 Toward Open Science

On June 2015, in his speech on "Open Innovation, Open Science, Open to the World", Carlos Moedas - Commissioner for Research, Science and Innovation – recognized that "there is a revolution happening in the way science works. Every part of the scientific method is becoming an open, collaborative and participative process" [11]. The term Open Science is widely used to refer this new vision of participatory scientific research. For example, the EGI community proposed the Open Science Commons as a new approach to digital research, summarizing the Open Science Commons Vision as "researchers from all disciplines have easy, integrated and open access to the advanced digital services, scientific instruments, data, knowledge and expertise they need to collaborate to achieve excellence in science, research and innovation" [12].

Recently, four top-level representatives of international science (the International Council for Science – ICSU, the InterAcademy Partnership – IAP, The World Academy of Sciences – TWAS and the International Social Science Council – ISSC) that are designed to represent the global scientific community in the international policy for science arena, developed an international accord on the values of open data in the emerging scientific culture of big data. The accord reminds that "openness and transparency have formed the bedrock on which the progress of science in the modern era has been based" and that "it is therefore essential that data that provide the evidence for published claims, the related metadata that permit their re-analysis and the codes used in essential computer manipulation of datasets, no matter how complex, are made concurrently open to scrutiny if the vital process of self-correction is to be maintained" [13].

The Open Science paradigm supports key aspects of the scientific method of investigation: openness, transparency, integrity and reproducibility. But, to realize its objectives, Open Science needs more than data sharing.

2.5.1 Open Knowledge

The effective (re-)use of data - especially when provided by different disciplinary infrastructures - requires the sharing of domain experts' knowledge. EGI referred to Knowledge as: "The human networks, understanding and material capturing skills and experience required to carry out open science" [12]. Experts' Knowledge stem from their education, culture, experience and is intertwined with the Community within which they work. Data is not knowledge, but expert's knowledge is essential to understand and use disciplinary data.

The term Open Knowledge is gaining importance, going over simple Open Data, referring to the open sharing – i.e. access, redistribution, reuse with no restriction – of any material including knowledge in any form. As the Open Knowledge International network says "Open knowledge is what open data becomes when it's useful, usable and used - not just that some data is open and can be freely used, but that it is useful – accessible, understandable, meaningful, and able to help someone solve a real problem" [14].

2.5.2 Virtual laboratories

Over the past decades several initiatives have started to support what is now the Open Science vision through information technologies. They brought to the building of digital infrastructures variously termed as Collaborative e-Research Communities, Collaborative Virtual Environments, Collaboratories, Science Gateways, Virtual Organisations, Virtual Research Communities, Cyberinfrastructures, Virtual Research Environments, Virtual Laboratories [15]. Although they are not synonyms, they share the idea of facilitating collaborative research at least in some aspect.

2.6 The concept of ECOPOTENTIAL Virtual Laboratory

The call does not specifically ask for something like a Virtual Laboratory. However, the ECOPOTENTIAL Consortium proposed the realization of an ECOPOTENTIAL Virtual Laboratory as the answer to the call request for "innovative solutions that will provide open and unrestricted access to interoperable ecosystem Earth Observation data and information". The Consortium recognized that the general objective of "supporting new measurements and observations, synthesis and interpretation of data for making all information and knowledge available to scientists, policy makers, citizens and other concerned stakeholders" can be achieved through the implementation of a Virtual Laboratory tailored to ecosystem science. To this aim we provide the following definition:

> The ECOPOTENTIAL Virtual Laboratory is a virtual environment supporting the activities of the ecosystem community-of-practice

This broad definition is based on attempts to clarify the different approaches to systems supporting collaborative science [16], with some modifications specifically related to the removal of references to specific requirements since they are the subject of a dedicated investigation in the project. In the definition above:

- *Virtual* means that there is not any necessity to physically centralize resources. For example, the VLab provides access to heterogeneous data, but data do not need to be moved from their original site.
- *Environment* means that the user experience is that of a controlled space where he/she can operate like in a physical laboratory.
- *Activities* refer potentially to the entire spectrum of actions, from the specifically scientific investigation to the communication and dissemination to stakeholders. They are clarified by the user requirement analysis and the range of supported activities may increase thanks to the extensibility of the platform.
- *Community-of-practice* refers to an informal group sharing interest on a topic. The definition of communities of practice has changed during years [17]; we adopt the definition by Wenger, McDermott and Snyder as "groups of people who share a concern, a set of problems, or a passion

about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis" [18].

3 ECOPOTENTIAL ARCHITECTURE PRINCIPLES

In the preparation of the DMP, ECOPOTENTIAL introduced a set of architecture and interoperability principles to facilitate data (and the associated software) discovery, access, (re-)use, and preservation:

AP1. To build the ECOPOTENTIAL data and services infrastructure on the existing and under development digital systems –noticeably, the digital systems managed by WP3, WP4, WP5, etc.

AP2. Not to impose any "common solution/specification" but advocate the use of open (international and Community) standards and interoperability APIs.

AP3. To provide a common, consistent, and "high-level" entry point ECOPOTENTIAL platform for discovering, accessing, and using ECOPOTENTIAL ecosystem services –for interoperability to GEOSS, Copernicus, and other EC-funded programmes.

AP4. To comply with the GEOSS Architecture Principles (see Annex C).

AP5. To comply with the GEOSS "resource sharing" and "resource management" principles –including quality and preservation (see Annexes A and B).

AP6. To comply with the EC Open Data Access principles (see Annex D and the Guidelines on Data Management in Horizon 2020, Version 1.0, 11 December 2013).

4 ECOPOTENTIAL DESIGN PRINCIPLES

The design principles translate the architecture principles in guidelines for the design of the ECOPOTENTIAL Virtual Laboratory platform.

4.1 Open Software Architectures

Taking into account that the effort dedicated to the realization of the Virtual Laboratory Platform is the 4.7% of the overall effort in the ECOPOTENTIAL project, the design bases on the Open Architecture paradigm in order to allow integration of existing mature solutions, minimizing the need of development from scratch.

The world of geospatial information is rapidly evolving with continuous provision of new tools, new data sources, new or revised specifications for data formats or service interfaces, new scenarios (such as recently *crowdsourcing*) and even completely new paradigms (like *open data* and *big data*). Therefore, ECOPOTENTIAL conceives a Virtual Laboratory as a member of a complex and evolving data and software ecosystem made of data sources, intermediate components and end-user applications. In particular, a VLab is an intermediate component that facilitates the connection between end-user applications and data sources, contributing to the software ecosystem evolution itself.

Living in an ever-changing context, the VLab must be also able to evolve in response to those changes. Indeed, although the VLab requirements can become clear during the course of the ECOPOTENTIAL project, in order to support the sustainability of outcomes, it is necessary to assure that the VLab architecture and implementation can (easily) evolve.

Software evolution has been the subject of several research works in the past (Table 1). A first classification [19] can be made between:

- *Centralized evolution*: where the pre- and/or post-deployment evolution is coordinated by a central authority
- *Decentralized evolution*: where the pre- and/or post-deployment evolution phases are based on activities of multiple teams

		When	
		Design-time (or pre-deployment) evolution	Post-deployment evolution
Who	Central authority (e.g., single vendor)	Design notations, methods, and tools; process systems; group communication and collaboration tools; configuration management	Release management systems; binary patch files; configurable distributed systems
	Decentralized group (e.g., multiple independent software vendors)	Same as above, with special support for loose coordination among geographically distributed team members (multiple sites or cross-organizational); open source	APIs, software plug-ins, scripting languages, open source, component architectures, and event-based systems

Table 1 Different categories of techniques to support software evolution

It is quite evident that a centralized evolution model is not an option for the ECOPOTENTIAL VLab for several reasons: a) a VLab is not fully based on software which is under control of a single organization (e.g. apps may be developed by external organizations); b) even the ECOPOTENTIAL Consortium as a whole does not control the full software suite (e.g. many components are open source and managed by a specific community); c) even assuming that the ECOPOTENTIAL Consortium could achieve the role of central authority, it exists only until the end of the project, whereas the sustainability of VLab must be considered also beyond the ECOPOTENTIAL project lifetime.

Decentralized software evolution can be achieved exposing the internal capabilities in any of multiple different ways: application programming interfaces (APIs), scripting languages, plug-ins, components architecture, event interface, source code. Each approach has its own advantages and drawbacks, and furthermore they are not mutually exclusive.

For the ECOPOTENTIAL purposes, the *source code* approach is not viable for several reasons: a) we cannot assume that all the components are or will be provided as open source; b) imposing the use of open sources would possibly exclude existing or future tools that could actually provide new functionalities (e.g. integration with big data platforms); c) imposing that evolution is based on collaborative working on open source would pose significant challenges in terms of *change analysis*, *fragility* and *composition*; d) the limited effort planned in ECOPOTENTIAL encourages to focus more on solutions that can be integrated in a loose way without requiring major development effort.

Likewise, *plug-ins*, *components architecture*, *event interface* approaches would need a major re-engineering of the existing tools which are not usually based on such approaches.

Instead, the provision of APIs is a loose approach which is provided by most of tools, and that can be easily enhanced through wrapping and extension. *Scripting language* is a possible complementary approach for implementing more complex functionalities.

Therefore we assume that the **ECOPOTENTIAL Virtual Laboratory adopt an Open Architecture with Decentralized Software Evolution based on APIs** allowing internal integration of existing tools and external interaction with other members of the geoinformation software ecosystem.

4.2 Brokered Systems of Systems

4.2.1 System of Systems Engineering

Interoperability is recognized as one of the main challenges for ECOPOTENTIAL. To address interoperability the ECOPOTENTIAL proposal is based on the successful experience of brokered architectures to implement Systems of Systems.

The notion of "System of Systems" (SoS) and "System of Systems Engineering" (SoSE) emerged in many fields of applications to address the common problem of integrating many independent, autonomous systems, frequently of large dimensions, in order to satisfy a global goal while keeping them autonomous. Therefore SoSs can be usefully described as follows: *systems of systems are large-scale integrated systems that are heterogeneous and consist of sub-systems that are independently operable on their own, but are networked together for a common goal* [20]. It is evident that this definition fits well in the ECOPOTENTIAL context where sub-systems like the INSPIRE infrastructure, Copernicus core and downstream services are clearly out of control of the ECOPOTENTIAL Consortium, and even from possible future exploitation scenarios.

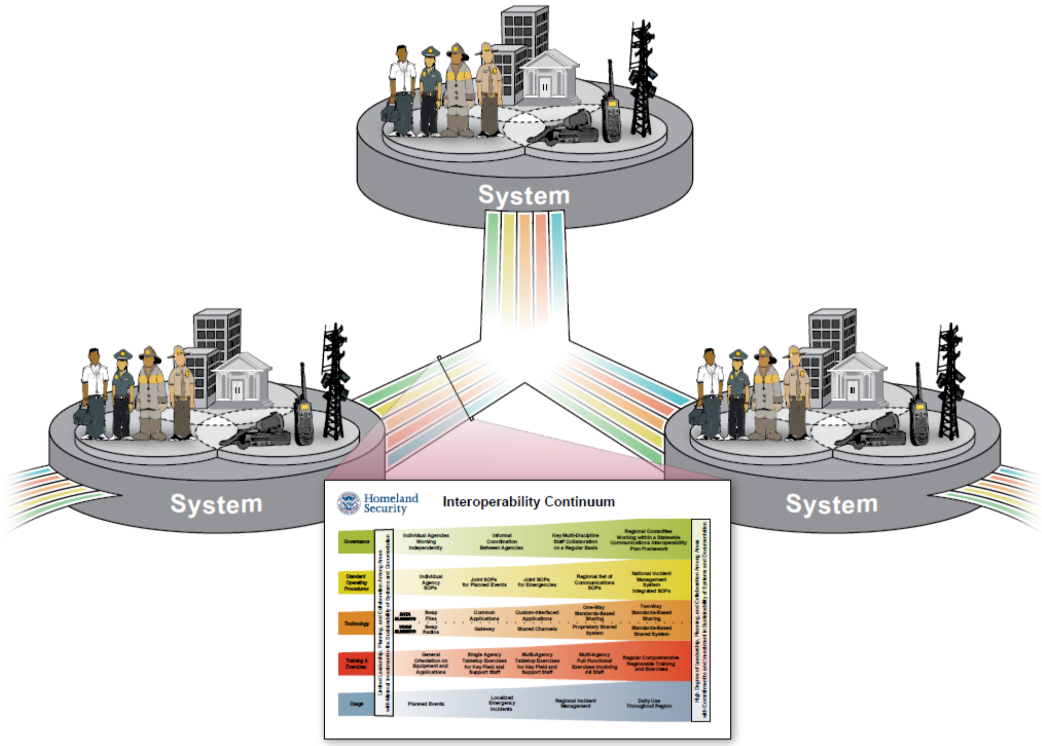

Figure 2 System of Systems in Practice – from [21]

4.2.2 Federation vs. Brokering

By a technical point-of-view, there are two general approaches for building a SoS: through *federation* and through *brokering*.

In the *federated approach*, a common set of specification (*federated model*) is agreed between the participating systems. It can range from a loose approach needing just the adoption of a suite of interface, metadata and data model standards to be applied by every participant, to a very strict approach imposing the adoption of the same software tools at every node. In every case, participants have to comply with the federated model (specifications or tools) and they need to make at least some change in their own systems. Therefore, this approach is feasible when:

a) the SoS governance has a strong mandate for imposing and enforcing the adoption of the federated model (e.g. as it happens with the INSPIRE Directive at the European level) to all the participants, or when the participants have a strong interest and commitment in participating in the SoS (as it happens in cohesive disciplinary communities)

b) the participant organizations have the expertise and skills for implementing the needed re-engineering of their own systems to make them compliant with the federated model

E-Commerce, e- Banking, and e-Government systems are typical examples where the federated approach fits well. In the geospatial world, the Open Geospatial Consortium (OGC) has been historically active in developing standard specifications, and the INSPIRE experience is an example where a central authority, the European Union, through a Directive, imposed a set of sharing principles, along with Implementing Rules, and Technical Guidelines, for establishing the Infrastructure for Spatial Information in Europe.

In the *brokered approach* [22] [23], no common model is defined, and participating systems can adopt or maintain their preferred interfaces, metadata and data models. Specific components (the *brokers*) are in charge of accessing the participant systems, providing all the required mediation and harmonization functionalities. The only interoperability agreement is the availability of documentation describing the published interfaces, metadata and data models. No (major) re-engineering of existing systems is required. This approach fits well in situations where the SoS governance does not have a specific mandate, and where the participant organization does not have a strong interest/commitment to be part of the SoS. In this case, third parties have the major interest in building the SoS. The brokered approach is also useful when the participant organization do not have the expertise for complying with complex specifications. This is a common situation in the Web world. In the geospatial world, the Global Earth Observation System of Systems (GEOSS) is the typical example of an overarching initiative where a third party, the Group on Earth Observation (GEO), has a specific interest in building a SoS collecting existing data systems with their own mandate and governance.

4.2.3 Standardization and brokering

Historically, in the geospatial world, federation has been the preferred approach. Initially, private companies, and research centers proposed their own technologies as the basis for a wide federation of data sources. Commercial tools are still widespread in GI systems for public authorities (e.g. Esri) and open source software suites are still the de-facto standards in some scientific communities (e.g. GSAC is UNAVCO's Geodesy Seamless Archive Centers software system for the geodesy community, THREDDS Data Server in the Meteo-Ocean community). Interoperability based on tool sharing has strong limitations, in particular due to adaptation to changes (e.g. centers using different versions of tools). In early 2000, such limitations pushed a more loosely-coupled approach based on standardization. The Open Geospatial Consortium (OGC) and ISO were and are particularly active in defining standards for geospatial data discovery and access. However, in parallel, many scientific and technological communities started their own standardization activities (e.g. TDWG in the biodiversity community). Although standardization allowed to mitigate many issues related to tools sharing, it demonstrated some shortcomings:

- *Slowness*: as a consensus-based approach “Standard development is a slow and difficult process” [24]. Standards react slowly to rapid changes in scenarios and requirements, in particular in presence of paradigmatic revolutions (e.g. Open Data movement, Big Data).
- *Complexity*: “Often the result can be large, complex specifications that attempt to satisfy everyone” [24]. Especially for interdisciplinary and multidisciplinary applications, the different requirements of heterogeneous communities would bring to very complex standards. For example: a standard suitable for Climate Change impact on biodiversity, should be able to support very specific requirements such as geological temporal scales (as required by the paleoclimate studies), species taxonomies (as required by ecological science) and so on.

Due to slowness and complexity of the standardization process, new standards are often developed by small groups, cohesive communities-of-practice (CoPs) and even companies, and once they become de-facto standards are then possibly approved by standardization bodies (as it happened with Google KML and UNIDATA netCDF in the OGC).

The resulting proliferation of standards posed clear interoperability issues. While some of them can be solved pushing the adoption of existing standards or accelerating the standardization process, others are not. In fact, many standards were born to answer to very specific requirements and to implement specific scenarios. **A single standard (or set of standards) would be either very complex – if it tries to accommodate all the heterogeneous requirements of geospatial applications from different communities – or underperforming for specific applications – if it tries to answer to a significant subset of requirements**.

A complex standard would pose severe barriers to implementation, requiring high IT expertise in interoperability which is usually not available by web developers, and often by data and research centers, or companies not specifically working on such topics. An underperforming standard would require communities to develop new standards or extend the existing ones for specific applications, quickly bringing again to standard proliferation and related interoperability issues.

A hybrid approach recently proposed and adopted (for example in the OGC) is based on modularity. Modular standards support basic and common requirements by default, and more specific requirements through dedicated modules. Although this approach reduces complexity, it poses interoperability issues related to different profiles (set of modules) implemented by different tools.

The brokered approach avoids those shortcomings, letting communities-of-practice free of defining their own specifications, and mediating between different specifications. Obviously mediation will happen at the lowest common level between specifications but it is generally sufficient for most interdisciplinary applications. Obviously brokering is not magic, the complexity of interoperability is still there. It is simply moved from data users and providers to the brokers. Data users and providers are set free from interoperability issues – i.e. they do not have to make their clients and server compliant with specifications anymore – but new components, the brokers, are in charge of handling all the complexity. However, this shift of complexity from clients/servers to brokers has two main advantages: (a) it implements the general engineering pattern called *separation-of-concerns*: where there is a specific functionality (interoperability), there should be a specific responsible (broker), (b) a third tier between clients and servers can host added-value services (e.g. semantics, data transformations). Obviously, brokered architectures present also possible issues, such as: (a) the middle-tier between clients and servers requires a specific governance, (b) as central architectural components, brokers may become single-points-of-failure, or bottlenecks. It is worth noting, that the former is currently addressed by the Brokering Governance WG[1] of the Research Data Alliance (RDA), and the latter can be solved resorting to specific architectural solutions based on redundancy and elastic computing.

Besides the previously described shortcomings, standards have an important benefit: the standardization process is the opportunity for requirements clarification, discussion and information modelling between experts. Therefore, although they cannot bring to a single standard for all the geospatial world, they help to avoid unnecessary proliferation of specifications, in particular without the needed quality. A brokered architecture could not manage thousands of (poorly designed) specifications. Therefore, when we talk about brokered approach we should actually consider a combined standardization+brokering approach. Standardization helps to reduce the redundant heterogeneity, while brokering addresses the remaining irreducible heterogeneity.

[1] https://rd-alliance.org/groups/brokering-governance.html

It is expected that different communities or sub-communities will develop standards building community federations, and then an overarching brokered System-of-Systems will integrate them enabling multidisciplinary applications.

In ECOPOTENTIAL, the choice of brokered architectures is fully justified by two main reasons:

a) There are several data sources of interest for ECOPOTENTIAL which are provided through heterogeneous protocols (interfaces, metadata and data models). In particular, many of them are not compliant with the widespread OGC standards. Just to mention some of them:
   a. The biodiversity community has defined its own set of specifications through the work of the Biodiversity Information Standards / Taxonomic Databases Working Group (TDWG)[2]
   b. In the meteo-ocean community, the UNIDATA THREDDS Data Server (TDS)[3] is widely adopted
   c. Many Open Data communities share the CKAN[4] technology for implementing data portals.
b) ECOPOTENTIAL has neither the mandate nor the capacity to impose and enforce standards or any federated model to the provider sub-systems.

4.2.4 Addressing interoperability through brokered architectures

The interoperability issue in the geospatial world can be summarized as the problem of allowing M different applications to interact with N different data sources: a MxN complexity problem (see Figure 3). By an architectural point-of-view, federated architectures can be implemented in a pure two-tier (client-server) environment. The M clients can interact with N servers in an easy way, because only one type of interaction is defined by the federated model. The MxN complexity is solved at client/server level changing both to make them compliant with the federation model. On the other hand, brokered architectures introduce a middle-tier between clients and servers, reducing the MxN potential interactions (each client needs to interact with any server) to M+N (each client and each server only need to interact with the brokers).

Since the connected sub-systems are and must be independently managed and autonomous, publishing functionalities are usually provided at local level according to the local policies. This means that federated/brokered services only include discovery and access and generally fruition services. ECOPOTENTIAL share this general approach: sub-systems are brokered with regards to access to resources ("read" mode), while any action causing modifications ("write" mode) is handled at sub-system level.

[2] http://www.tdwg.org
[3] www.unidata.ucar.edu/software/thredds/current/tds/
[4] http://www.ckan.org

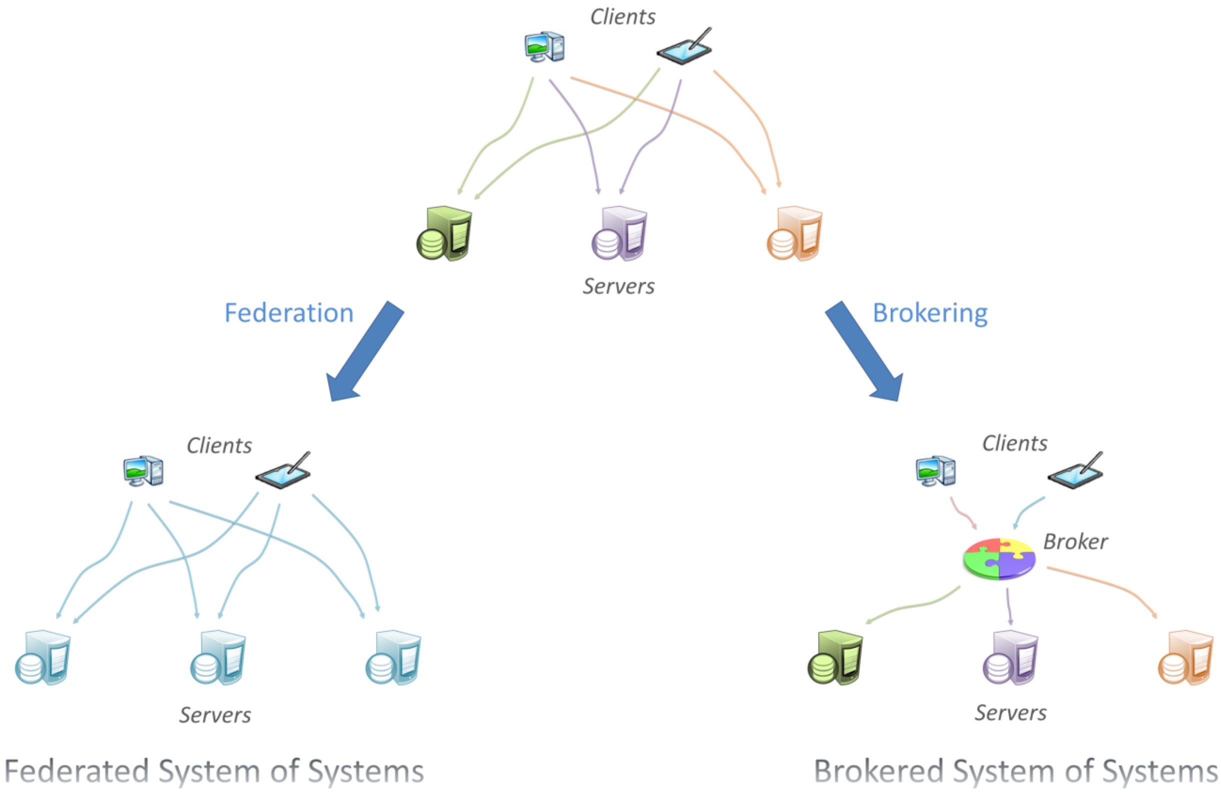

Figure 3 Federated vs. Brokered Architectures for Systems of Systems

4.3 ECOPOTENTIAL service provision model

Over the past years, the evolution of Information Technologies, allowing ubiquitous connectivity, imposed the *cloud computing* paradigm. Cloud computing can be defined as "*a model for enabling ubiquitous, convenient, on-demand network access to a 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*" [25].

The cloud model includes three different kinds of services [25]:

- *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. Examples are Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3).
- *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. Examples are Google App Engine and Microsoft Azure.
- *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. Examples are Google Docs, or Microsoft Office Online.

The cloud model is particularly appealing for the provision of services. Indeed, it presents some advantages: a) it widen the range of users, requiring only a browser and a good connectivity which is currently easy to achieve even in mobility, b) it separates responsibilities, delegating support services (hardware and software management, accounting and billing) to cloud providers, and allowing developers to focus on their own application.

In ECOPOTENTIAL, where there is no particular need for a different approach, **applications will be provided as SaaS to end-users**. This means that end-users will be able to use the applications simply accessing the Virtual Laboratory with their own browser.

The VLab App Developers will interact with the VLab according to SaaS and PaaS model. The **VLab PaaS will provide the APIs and the programming environment for fast development and deployment of applications**. The VLab SaaS may also provide the developers with ancillary services, for example to access documentation, to communicate with the VLab Administrator, or with other VLab App Developers (e.g. forum, chat).

The VLab platform, composed of PaaS for developers, and SaaS for users in general, will be designed to be deployed either on proprietary infrastructure or on cloud IaaS.

4.4 Orthogonality of resource-sharing and security architectures

ECOPOTENTIAL requirements can be broadly classified into two categories:

- Resource-sharing requirements, expressing needs for assuring seamless sharing of open geospatial resources
- Security requirements, expressing the needs for identifying users, checking authorizations, logging activities

The general ECOPOTENTIAL architecture can be decomposed in a Resource-sharing Architecture describing the structure and interaction of components fulfilling resource-sharing requirements, and a Security Architecture describing the structure and interaction of components fulfilling security requirements. In ECOPOTENTIAL we assume the *orthogonality* of the two architectures, meaning that any change in one of them should not affect the other one. This is a common assumption in software architectures and it strictly derives from the orthogonality (independence) of resource-sharing and security requirements. The advantage of orthogonality is that it allows decomposing architectures handling each aspect separately.

4.5 ECOPOTENTIAL design principles

It is possible to summarize the outcomes of discussions above in the following architectural principles:

DP1. ECOPOTENTIAL Virtual Laboratory adopts an Open Software Architecture
DP2. ECOPOTENTIAL Virtual Laboratory is developed integrating and adapting existing software solutions
DP3. ECOPOTENTIAL Virtual Laboratory adopts a Decentralized Software Evolution
DP4. ECOPOTENTIAL Virtual Laboratory is made of software components interacting through (low-level) APIs
DP5. ECOPOTENTIAL Virtual Laboratory is the common infrastructure of a brokered System of Systems
DP6. ECOPOTENTIAL Virtual Laboratory exposes a set of (high-level) APIs for interaction with the external environment
DP7. ECOPOTENTIAL Virtual Laboratory is accessible according to the Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) models, for end-users and developers respectively
DP8. ECOPOTENTIAL Virtual Laboratory can be deployed either on private infrastructures or commercial or public clouds providing Infrastructure-as-a-Service (IaaS) capabilities.
DP9. ECOPOTENTIAL Virtual Laboratory security architecture is orthogonal to the ECOPOTENTIAL Virtual Laboratory resource-sharing architecture.

5 ECOPOTENTIAL SYSTEM ARCHITECTURE OVERVIEW

5.1 Architecture description

A system architecture is the set of "fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution" [26]. An architecture is described through an architecture description which is "a set of products that documents an architecture in a way its stakeholders can understand and demonstrates that the architecture has met their concerns" [27].

A complex system cannot be effectively described through a single over-compassing description. It should provide a lot of information ranging from high-level aspects like stakeholders' interactions with the system, to very low-level aspects such as software objects methods, interfaces and technological choices. Different stakeholders would find most of the information unnecessary and too detailed for those aspects they are not specifically interested in. *Viewpoint modelling* addresses this issue providing different views of the same architecture. "*A view is a representation of one or more structural aspects of an architecture that illustrates how the architecture addresses one or more concerns held by one or more of its stakeholders*" [27].

The following paragraphs provide the ECOPOTENTIAL Virtual Laboratory description according to the following main views adopted in the ISO Reference Model for Open Distributed Processing (RM-ODP) [28]:

- Enterprise Viewpoint
- Computational Viewpoint
- Information Viewpoint
- Engineering Viewpoint
- Technology Viewpoint

5.2 Enterprise Viewpoint

The enterprise viewpoint […] is concerned with the purpose, scope and policies governing the activities of the specified system within the organization of which it is a part;
[28]

The enterprise viewpoint focuses on the actors, their interactions in scenarios, use-cases and it allows the elicitation of user requirements and then system requirements.

5.2.1 Actors

ECOPOTENTIAL identifies a set of Actors, which is a set of user categories involved in: a) the setup and operation of the Virtual Laboratory, b) the use of Virtual Laboratory resources, and finally, c) the use of applications based on the Virtual Laboratory. They are

Actor	Acronym	Description
Virtual Laboratory Provider	VLab Provider	The VLab Provider is the person/organization that provides the VLab capacities.
Virtual Laboratory Administrator	VLab Admin	The VLab Admin is the person who manages a Virtual Laboratory configuring it for VLab users and providing support.
Virtual Laboratory End User	VLab End User	The VLab End User is a member of the Ecosystem CoP who makes use of the VLab capabilities

Virtual Laboratory App Developer	VLab App Developer	The VLab App Developer is an intermediate user, a person who develops and manages applications based on the VLab APIs.
Virtual Laboratory Consumer	VLab Consumer	A VLab Consumer is a person who makes use of VLab capabilities, which is either a VLab End User or a VLab App Developer.

Table 2 Description of ECOPOTENTIAL Virtual Laboratory actors

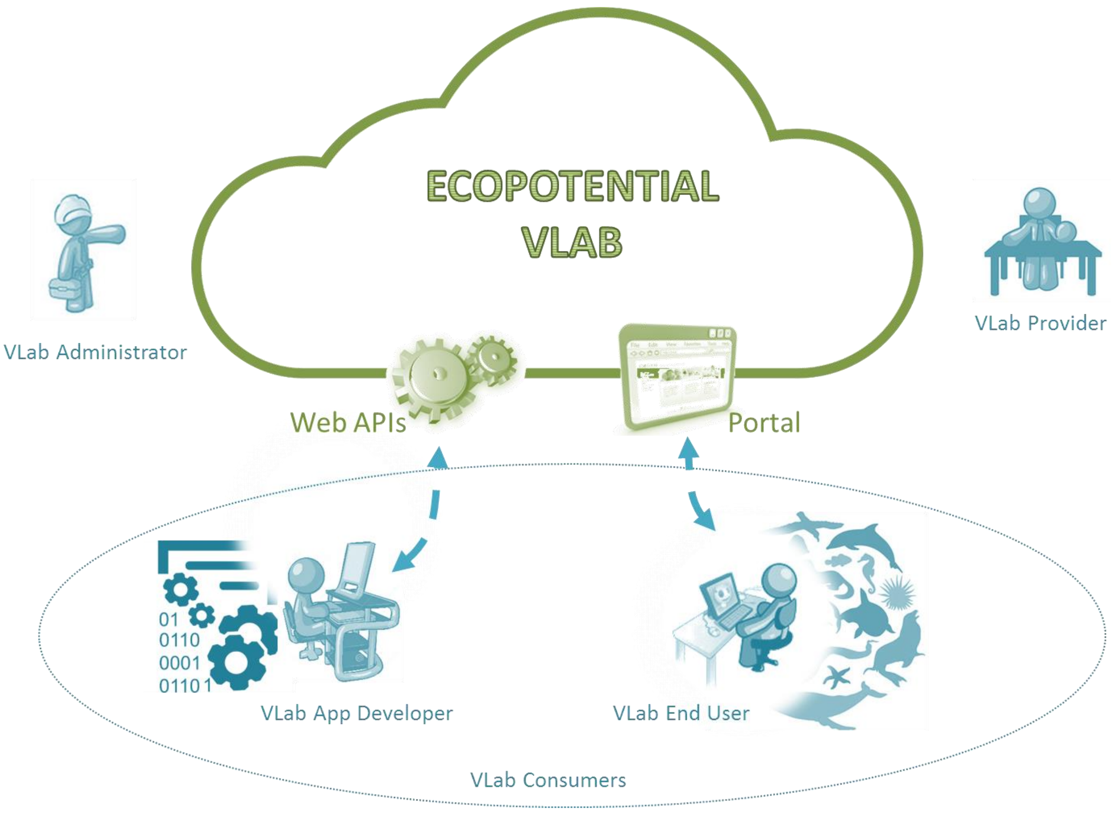

Figure 4 Virtual Laboratory actors (in green)

5.2.2 User scenarios and requirements

ECOPOTENTIAL User Requirements are collected from different sources:

a) Call text [2]
b) ECOPOTENTIAL DoW [29]
c) Collection of user requirements from ECOPOTENTIAL WP7 "Ecosystem Services", WP8 "Cross-scale interaction", WP9 "Requirements of future protected areas"
d) Previous work in relevant initiatives and programmes at national, regional, European and international level (including Copernicus, INSPIRE, GEOSS)

Sources a) and b) assure the expected impact and compliance with the project agreements. Source c) provides information on user needs. Source d) assures that the project outcomes are in line with the major initiatives in the sector.

At the stage of the preparation of this release of the deliverable (May 2016), according to the DoW, no formal outcome was expected from source c), therefore the current status of user needs is based on sources a), b) and d).

In terms of user requirements, the ECOPOTENTIAL Virtual Laboratory was conceived as a typical resource sharing system with a specific focus on solving interoperability issues to facilitate open knowledge scenario. The high-level use-cases are those needed to support the typical resource sharing scenario shown in Figure 5, including Publishing (supporting upload of relevant resources), Discovery (supporting search for relevant resources), Evaluation (supporting inspection of resources to evaluate their value and relevance), Access (supporting retrieval of relevant resources), Use (from simple visualization to complex processing where required). It is represented as a cycle because the result of resource usage may be a new resource to be published. The figure also shows a Management use case which underpins all the information life-cycle.

Due to the need of sharing heterogeneous resources within the project and with the outside world, a specific attention on interoperability issues is required.

Figure 5 The typical high-level scenario in the resource sharing

In ECOPOTENTIAL the term *resource* encompasses:

- Data
- Semantic assets
- Scientific workflows
- Analytic services

Table 3 shows the main user scenarios for the ECOPOTENTIAL VLab.

User Scenario	Description

S1.	Search for datasets	The VLab End User: a) searches for available datasets per geographical coverage (including Protected Area identifier), temporal extent, keywords, concepts; b) evaluates available datasets through metadata; c) downloads relevant datasets in the preferred format, resolution, etc.
S2.	Search for Protected Areas	The VLab End User: a) browses Protected Areas; b) chooses one Protected Area; c) gets available information about the selected Protected Area (including available Storylines)
S3.	Search for Storylines	The VLab End User: a) browses Storylines; b) chooses one Storyline; c) gets available information about the selected Storyline (including available Protected Areas and Workflows)
S4.	Publish resources	The VLab End User publishes a resource which can be: i) an existing data system; ii) a set of resource artifacts previously unpublished (or to be mirrored)
S5.	Add a Protected Area	The VLab End User add a Protected Area with all the required information to the VLab
S6.	Add a Storyline	The VLab End User add a Storyline with all the required information to the VLab including which ECOPOTENTIAL Protected Areas it refers to
S7.	Add a Workflow	The VLab End User add a Workflow with all the required information to theVLab including which ECOPOTENTIAL Storylines it refers to, optionally uploading the source/executable code and a web service
S8.	Run a Workflow for a Protected Area	The VLab End User: a) browses Protected Areas; b) chooses one Protected Area; c) browses the available Storylines for that Protected Area; d) chooses a Storyline; e) browses the available Workflows; f) chooses one Workflow; g) select input datasets available for that Workflow on that Protected Area; h) runs the Workflow; i) accesses the result
S9.	Run a Workflow for a Storyline	The VLab End User: a) browses Storylines; b) chooses a Storyline; c) browses the available Workflows; d) chooses one Workflow; e) browses supported Protected Areas; f) chooses one Protected Area; g) select input datasets available for that Workflow on that Protected Area; h) runs the Workflow; i) accesses the result
S10.	Run apps	The VLab End User: a) browses the marketplace searching for apps; b) chooses one app; c) downloads/access the app; d) runs the app. Steps a)-c) are needed only the first time
S11.	Develop a new app	The VLab App Developer: a) accesses API documentation; b) downloads the Javascript/HTML5/CSS library, if needed; c) develops the app in his/her preferred environment; d) publishes the app in the ECOPOTENTIAL marketplace
S12.	Annotate a resource	The Vlab End User can annotate a resource (dataset, Protected Area, Storyline, Workflow) providing a quality feedback

Table 3 User scenarios for the ECOPOTENTIAL Virtual Laboratory

Table 4 summarizes the main user requirements obtained as elicitation from user scenarios, and general requirements.

User Requirement	Description
Data publishing	The VLab Consumer is able to publish single datasets or connect data systems with minimal interoperability agreements
Harmonized access to data	The VLab Consumer is able to seamlessly discover and download data from heterogeneous sources

Data harmonization	The VLab consumer is able to download data harmonized in terms of format, spatial and temporal coverage, coordinate reference system, resolution.
Scientific workflow publishing	The VLab Consumer is able to publish a formal representation of a scientific model
Scientific workflow access	The VLab Consumer is able to discover, visualize and run scientific models
Analytic service publishing	The VLab Consumer is able to publish a Web service implementing a scientific model
Analytic service run	The VLab Consumer is able to run a Web service implementing a scientific model
Semantic enrichment	The VLab Consumer is able to use semantic assets for suggestions, multilingual discovery, etc.
User feedback	The VLab Consumer is able to provide feedback and annotations on resources
User profile support	The functionalities offered to VLab actors are based on their profile
Relocation of platform	The VLab Administrator is able to move the platform to a different VLab Provider for performance or financial considerations

Table 4 User requirements for the ECOPOTENTIAL Virtual Laboratory

5.2.3 Constraints and assumptions

The main constraint for ECOPOTENTIAL Virtual Laboratory Platform is that the resources must comply with the ECOPOTENTIAL Data Management Plan [30] to be available in the VLab.

5.2.4 System Requirements

ECOPOTENTIAL System Requirements are collected from different sources:

a) Call text [2]
b) ECOPOTENTIAL DoW [29]
c) Elicitation from user requirements (see Section §5.2.1)
d) Data Management Plan [30]
e) Specific requirements from ECOPOTENTIAL WP3 "Earth Observation Data and Processes Infrastructure", WP4 "Earth Observation Data Generation and Harmonization", WP5 "In situ Monitoring Data", WP6 "EO-based Ecosystem Modelling"

Table 5 reports the identified system requirements. They are classified in functional requirements (describing *what* the system has to provide), and non-functional requirements (describing *how* the system has to provide functionalities).

Code	Name	Description
FR1	Dataset discovery	The system provides discovery of datasets based on different criteria including at least: a) geographical coverage expressed as bounding box; b) temporal extent expressed as start and end date/hour; c) keywords present in multiple metadata fields; d) data provider expressed as catalog/inventory name;
FR1.1	Dataset discovery protocols (data sources)	The system supports the data discovery protocols identified in the DMP for connecting data sources (see section §5.4.3)
FR1.2	Dataset discovery protocols (clients)	The system publishes the data discovery protocols identified in the DMP for communication with clients (see section §4.4.1): At the minimum the following discovery protocols will be supported:

		a) OpenSearch (and relevant extensions) b) OGC CSW 2.0 ISO Profile
FR2	Semantic discovery	The system provides semantic enhancements for discovery, supporting multilingualism, suggestions, and search for related terms.
FR2.1	Semantic discovery protocols	The system provides the possibility to connect to SKOS RDF knowledge bases publishing a SPARQL interface.
FR2.2	Semantic discovery – knowledge bases	The system is able to access at least the GEMET (GEneral Multilingual Environmental Thesaurus) thesaurus for supporting multilingual discovery.
FR3	Dataset access	The system provides access to datasets from heterogeneous data systems
FR3.1	Dataset access protocols (data sources)	The system supports the data access protocols identified in the DMP for connecting data sources (see section §5.4.3)
FR3.2	Dataset access protocols (clients)	The system publishes the data access protocols identified in the DMP for communication with clients (see section §4.4.1). At the minimum data can be accessed through any of the following protocols: a) OGC WCS, b) OGC WFS, c) OGC WMS,
FR3.3	Dataset access formats (data sources)	The system supports the data formats identified in the DMP for accessing data sources (see section §4.4.1)
FR3.4	Dataset access formats (clients)	The system supports the data formats identified in the DMP for communication with clients (see section §4.4.1)
FR4	Dataset transformation	Through the system, a user can access datasets from different data sources and retrieve them on a Common Grid Environment (same resolution, same CRS, same format, etc.). The system supports basic data transformation functionalities including: a) subsetting b) interpolation c) reprojection on multiple Coordinate Reference Systems d) data format transformation
FR5	Algorithm discovery	The system provides discovery of algorithms based on keywords and description content
FR6	Algorithms access	The system provides access to the code implementing the algorithm
FR7	Scientific workflow discovery	The system provides discovery of scientific workflows based on different criteria including at least: a) Protected area b) Storyline
FR8	Scientific workflow visualisation	The system provides a graphic visualization of a scientific workflow
FR9	Scientific workflow invocation	The system allows to run a scientific workflow on selected datasets
FR10	AAA	The system must support Authentication, Authorization and Accounting allowing collecting information about the use for both technical and marketing purposes.
FR11	Data Publishing	The system support resource publishing on a long-term preservation system, making the resource available for discovery and use
FR12	Resource Annotation	The system supports annotation of resources
FR13	Data registration in GEOSS	Data available in the Vlab are accessible also through GEOSS (related to FR1.2 and FR3.2)

NFR1	Seamless discovery and access	The system provides discovery and access of heterogeneous resources through any of the available protocols
NFR2	APIs	The system functionalities are accessible both server-side (for integration of tools enhancing system capabilities) and client-side (for application development through mash-up) through APIs
NFR2.1	APIs implementation	The system supports at least: a) server-side open interface b) Web APIs (HTML5-JavaScript-CSS library)
NFR3	Availability	The system must assure high availability
NFR4	Performance	The system must assure adequate performances
NFR5	Scalability	The system must assure adequate scalability in terms of number of data sources, number of users, number of requests, etc.
NFR6	Security	The system must assure security
NFR7	Usability	The system must be user-friendly for both end-users and application developers
NFR8	Extensibility	The system must be extensible to support new data sources protocols, new apps without major changes
NFR9	Accuracy	The system should not introduce loss of data quality (e.g. in data transformations)

Table 5 ECOPOTENTIAL system requirements

5.3 Computational Viewpoint

Computational VP is concerned with the functional decomposition of the system into a set of objects that interact at interfaces - enabling system distribution.
[28]

Figure 6 shows the high-level architecture of the ECOPOTENTIAL Virtual Laboratory platform. It includes the following layers:

- **Resource Access** layer: this layer provides functionalities for publishing, discovery and access resources on heterogeneous data systems. The figure shows the main functionalities provided by this layer: metadata QA/QC, harmonization, discovery, access.
- **Workflow Access** layer: this layer provides functionalities supporting workflows based on the harmonized resources provided by the lower layer. The figure shows the main functionalities provided by this layer: discovery, access, invocation and management of worflows.
- **User Interface** layer: this layer provides user-friendly access to resources for end-user. The access to the lower layers is provided by open APIs allowing intermediate users (e.g. app developers) to create new applications. The figure shows the different intrfaces provided by this layer: portal, apps, marketplace for different kind of resources.

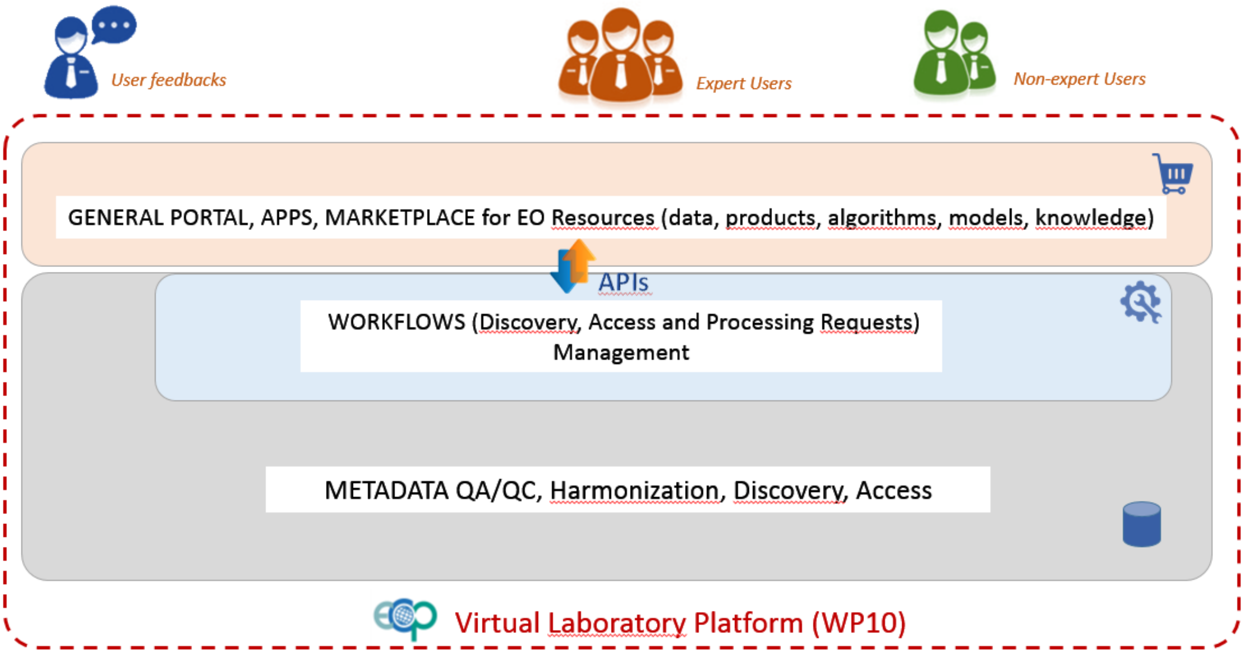

Figure 6 ECOPOTENTIAL layered architecture

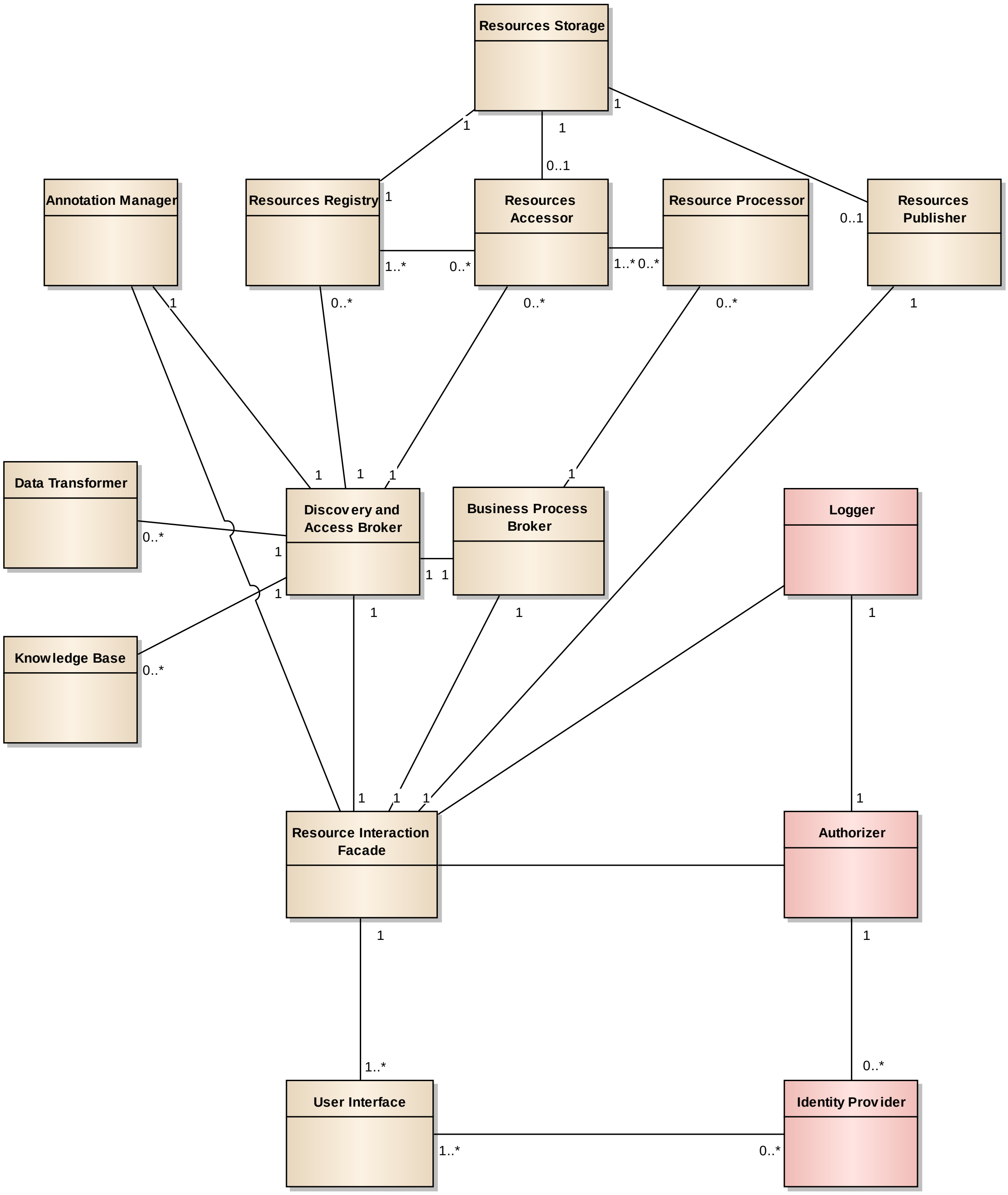

Figure 7 Main components of the ECOPOTENTIAL logical architecture (security components in red)

Figure 7 shows the UML class diagram of the main functional components in the ECOPOTENTIAL architecture. Components in red are not involved in the resource-sharing functionalities of the ECOPOTENTIAL system; they are either components of the security architecture or ancillary components improving the ECOPOTENTIAL overall system capabilities.

The main functional components are described in Table 6 along with a reference to the functional and non-functional requirements they contribute to fulfil.

Component	Description	Relevant requirements/constraints
Resource Storage	It hosts representations of resources in a specific source	FR11
Resources Registry	It provides discovery of resources from a specific source	FR1, FR2, FR5, FR7
Resources Accessor	It provides access to resources from a specific source	FR3, FR6, FR8
Resource Publisher	It provides upload of resources on (one or more) sources connected to the Virtual Laboratory platform	FR11
Resource Processor	It provides processing of resources	FR9
Annotation Manager	It allows publishing, discovery, and access of annotations to resources.	FR12
Discovery and Access Broker	It accesses multiple Resources Registries and Resources Accessors providing harmonized discovery and access to heterogeneous sources. It also merges providers metadata from registries with users metadata from annotations	FR1, FR2, FR3, FR4, FR5, FR6, FR7, FR13 NFR1, NFR8
Business Process Broker	It accesses multiple Resource Processors providing harmonized access. It also provides automated adaption of inputs through the Discovery and Access Broker.	FR9
Data Transformer	It transforms data changing resolution, Coordinate Reference System, format, etc on-the-fly. The content and semantic level of data is not changed.	FR4
Knowledge Base	It provides encoding of knowledge, to support advanced discovery services	FR2
Resource Interaction Facade	It provides a common and simplified interface to the VLab services, simplifying the application development	NFR2
Authorizer	It is the policy decision point checking if the user is authorized to perform an operation based on his/her identity and permissions	FR10 NFR6
Identity Provider	It checks the user's identity	FR10 NFR6
Logger	It stores information about the status of the data sources, and users' activities, for logging, accounting and monitoring purposes. In particular request and response will be monitored and evaluated.	FR10 NFR6
User Interface	It handles the interaction between the user and the system. It includes GUIs allowing presentation of maps with pan and zoom, layer selection. It must support 2D maps and 3D landscape scenes. It must provide data and metadata s tables and charts.	NFR7

Table 6 ECOPOTENTIAL main components

From Table 6 it is evident the core roles of the Discovery and Access Broker – harmonizing interaction with multiple resources - which impacts on many functional and non-functional requirements. Actually, as a computational architecture, it addresses all the functional requirements and also impacts on some non-functional requirements. However, most of the non-functional requirements are addressed by the distribution architecture discussed in the Engineering Viewpoint in section 5.5, and by the implementation and deployment choices described in sections 6 and 7 in particular through the Infrastructure-as-a-Service deployment.

5.4 Information Viewpoint

Information VP is concerned with the kinds of information handled by the system and constraints on the use and interpretation of that information. [28]

5.4.1 The ECOPOTENTIAL information model

As discussed in the user scenarios, ECOPOTENTIAL introduces and makes use of some relevant concepts. In particular, users will interact with the VLab in terms of Protected Areas, Storylines and/or Workflows. They will be the link with other concepts like Data, Algorithms, and Services which may remain hidden by the VLab technical implementation. Figure 8 shows the relationships between the main high-level concepts in ECOPOTENTIAL.

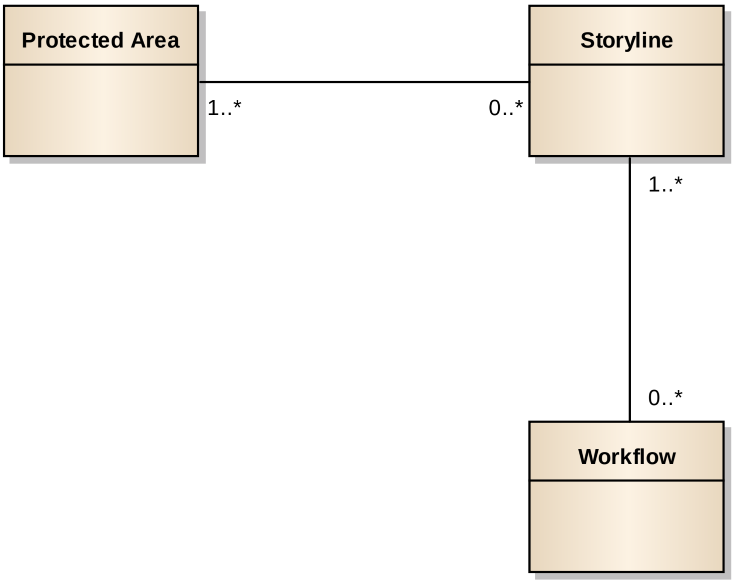

Figure 8 High-level concepts in ECOPOTENTIAL

5.4.2 Resource sharing in ECOPOTENTIAL

As a project finalized to the creation of a Virtual Laboratory facilitating the workflows based on heterogeneous resources, the characteristics of information handled and shared by the system is a fundamental aspect.

ECOPOTENTIAL addresses two main challenges concerning information handled by the Virtual Laboratory:

- *Heterogeneity*: the connected data sources vary largely in terms of service interfaces, metadata and resource models;
- *Semantics*: the content can be annotated and interpreted according to different semantics.

5.4.3 Heterogeneity

The ECOPOTENTIAL Virtual Laboratory aims to facilitate the use of many different kinds of geospatial resources (see Figure 9), including:

- User feedback
- GEOSS resources
- Satellite data
- In-situ data
- Processing Algorithms
- Models/Workflows
- Model results (products)
- Infrastructures

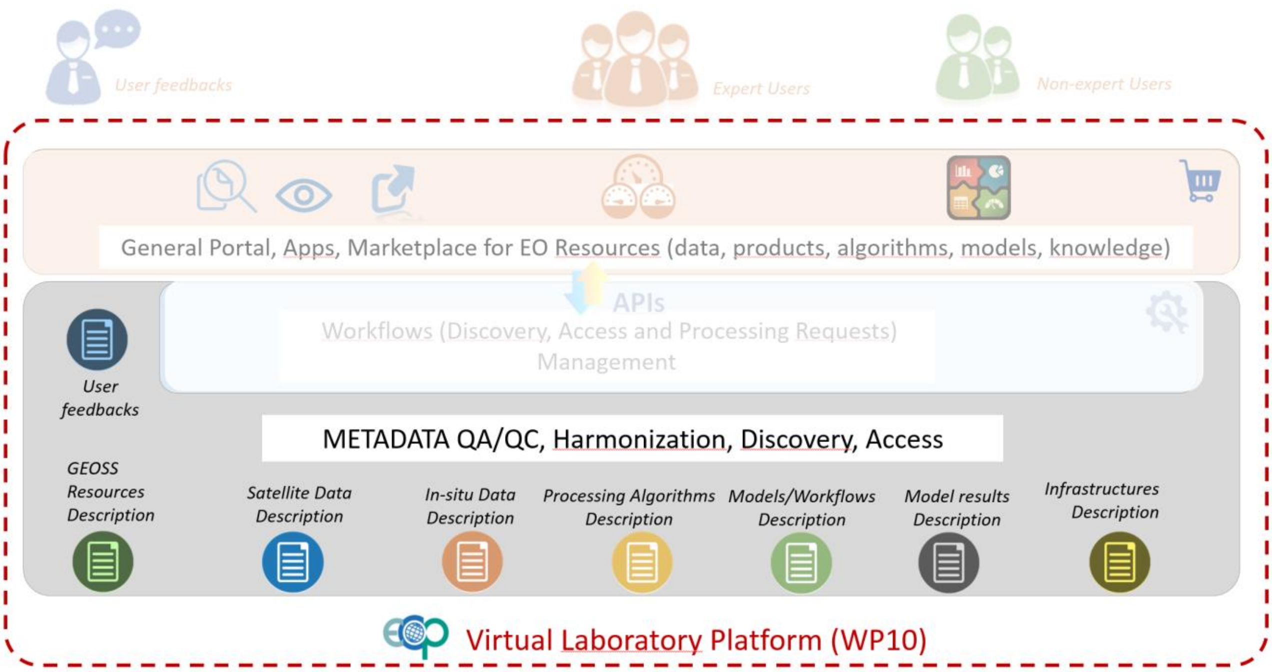

Figure 9 ECOPOTENTIAL resources

For some resources like workflows which are not widely shared yet, it is possible to define a common model to be used in ECOPOTENTIAL, but for data this is not possible. Geospatial data comes in many different shapes (see section §2.4), and ECOPOTENTIAL cannot assume any kind of standardization, as described and discussed in the ECOPOTENTIAL DMP [30]. As such the VLab platform must take care of all the mediation, harmonization and transformation actions needed to make geospatial data easily discoverable, accessible, and usable.

It is worth noting that heterogeneity affects all the information lifecycle in ECOPOTENTIAL. *Figure 10* shows the heterogeneity of discovery and access protocols limiting to those identified through a first survey.

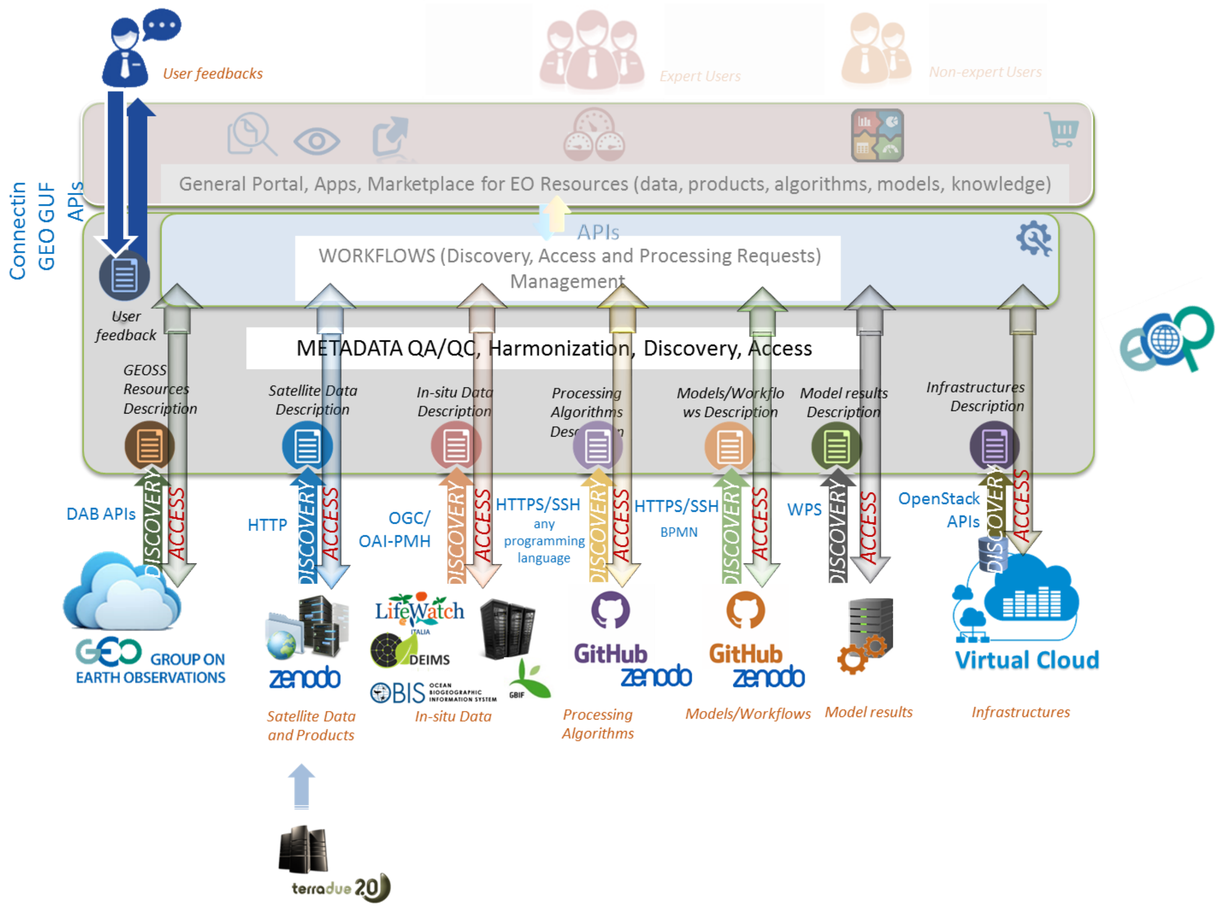

Figure 10 Heterogenity of discovery and access protocols in ECOPOTENTIAL

In general, the Virtual Laboratory must be able to handle different service interfaces and metadata/data models for data discovery and access. The ECOPOTENTIAL DMP aims to provide the full list of data and data service specifications to be supported. Table 7 shows protocol requirements to discover, access and invoke ECOPOTENTIAL resources, based on DMP and preliminary investigations:

Resource category	Resource type	Resource system	Discovery interface	Metadata model/encoding	Access interface	Data model/encoding
Satellite data	Landsat	Zenodo[1]	Metadata harvesting	MTL	Direct download	GeoTIFF
	MODIS	Zenodo[1]	Metadata harvesting	HDF-EOS	Direct download	HDF
	Sentinel-1	Zenodo[1]	Metadata harvesting	XML manifest file	Direct download	GeoTIFF
	Sentinel-2	Zenodo[1]	Metadata harvesting	XML manifest file	Direct download	JPEG2000
	Sentinel-3	Zenodo[1]	Metadata harvesting	CDL	Direct download	netCDF
Satellite derived products	water turbidity maps	Zenodo	Metadata harvesting	INSPIRE+Copernicus Land Service	Direct download	To be defined
	Land Cover/use (LCLU) and LCLU change maps	Zenodo	Metadata harvesting	OGC, ISO	Direct download	KEA, GeoTIFF

	Habitat and habitat change maps	Zenodo	Metadata harvesting	*Not available*	Direct download	KEA, GeoTIFF, NATURA2000 standard data format
	Soil moisture	Zenodo	Metadata harvesting	*Not available*	Direct download	GeoTIFF
	Seasonal Water Bodies	Zenodo	Metadata harvesting	Copernicus Land Service	Direct download	GeoTIFF
	Snow Cover Area	Zenodo	Metadata harvesting	*Not available*	Direct download	NOAA standards (OGC, netCDF-CF, THREDDS, and OPeNDAP) and GeoTIFF
	Snow Cover Maps	EURAC	OGC CSW	INSPIRE	WMS	GeoTIFF
	Water Bodies Delineation	Zenodo	Metadata harvesting	Copernicus Land Service	Direct download	GeoTIFF
	Maps of landscape measures indicating fragmentation or connectivity	Zenodo	Metadata harvesting	Copernicus Land Service	Direct download	KEA, GeoTIFF
In-situ data	In-situ existing datasets	LTER	DEIMS	INSPIRE, EML		OGC O&M
	In-situ research locations	LTER	*Not available*	INSPIRE EF Data Model (Environmental Monitoring Facility)	*Not available*	*Not available*
	In-situ generated datasets (file-based)	Zenodo	Metadata harvesting	INSPIRE, EML	Direct download	
	In-situ generated datasets (spatial data)	*Not available*	Metadata harvesting	INSPIRE, EML	OGC WMS, WCS or WFS	*Not available*
	In-situ generated datasets (sensor data)	*Not available*	Metadata harvesting	INSPIRE, EML	OGC SOS	*Not available*
	LifeWatch-ITA Data Products	LifeWatch-Italy Data Portal	*Not available*	LTER DEIMS	direct download	*Not available*
Model results	Model results	*Not available*	*Not available*	*Not available*	*Not available*	*Not available*
Indicators/Indices	Indicators/Indices	*Not available*	*Not available*	*Not available*	*Not available*	*Not available*
Processing algorithms		Zenodo	Metadata harvesting	*To be defined*	Direct download	Various
Models/Workflows description		Zenodo	Metadata harvesting	*To be defined*	Direct download	BPMN

Processing infrastructures	Ecosystem models	CNR	Metadata harvesting	WPS GetCapabilities	WPS	Various
Processing Infrastructures	Virtual Cloud	Various	OpenStack API	*To be defined*	OpenStac k API	Various

Table 7 Protocol requirements to discover, access and invoke ECOPOTENTIAL resources

[1] Relevant datasets will be collected and mirrored on Zenodo

5.4.4 Semantics

The Virtual Laboratory addresses semantics through a query expansion strategy. When a query is submitted to the VLab, the VLab can ask external semantics services, to resolve keywords, providing “related” terms back. The returned concepts are used as keywords of multiple geospatial queries [31]. Then, the results from geospatial queries include responses not only to the original keywords but also to semantically related terms. (See Knowledge Base component in Figure 7, and Figure 11 in section §5.5, below.)

The use of external semantic services enables extensibility. The type of relationships that can be used depends on the underlying knowledge bases. For example, SKOS (Simple Knowledge Organization System) provides a standard way to represent knowledge organization systems using the Resource Description Framework (RDF), allowing to express basic relationships such as “broader”, “narrower”, etc. supporting the encoding of thesauri, classification schemes, subject heading lists and taxonomies.

The query expansion strategy enables multilingual queries. Indeed, if one of the knowledge bases includes translations as “related” terms (e.g. the General Multilingual Environmental Thesaurus: GEMET) , the system will send different queries for each translation. Therefore, the query will return datasets whose description include either the proposed keyword or any of its supported translations. This is extremely important whenever there is not any obligation to compile metadata in a specific language.

5.5 Engineering Viewpoint

Engineering VP is concerned with the infrastructure required to support system distribution.
[28]

Figure 11 shows the engineering view of the ECOPOTENTIAL architecture.

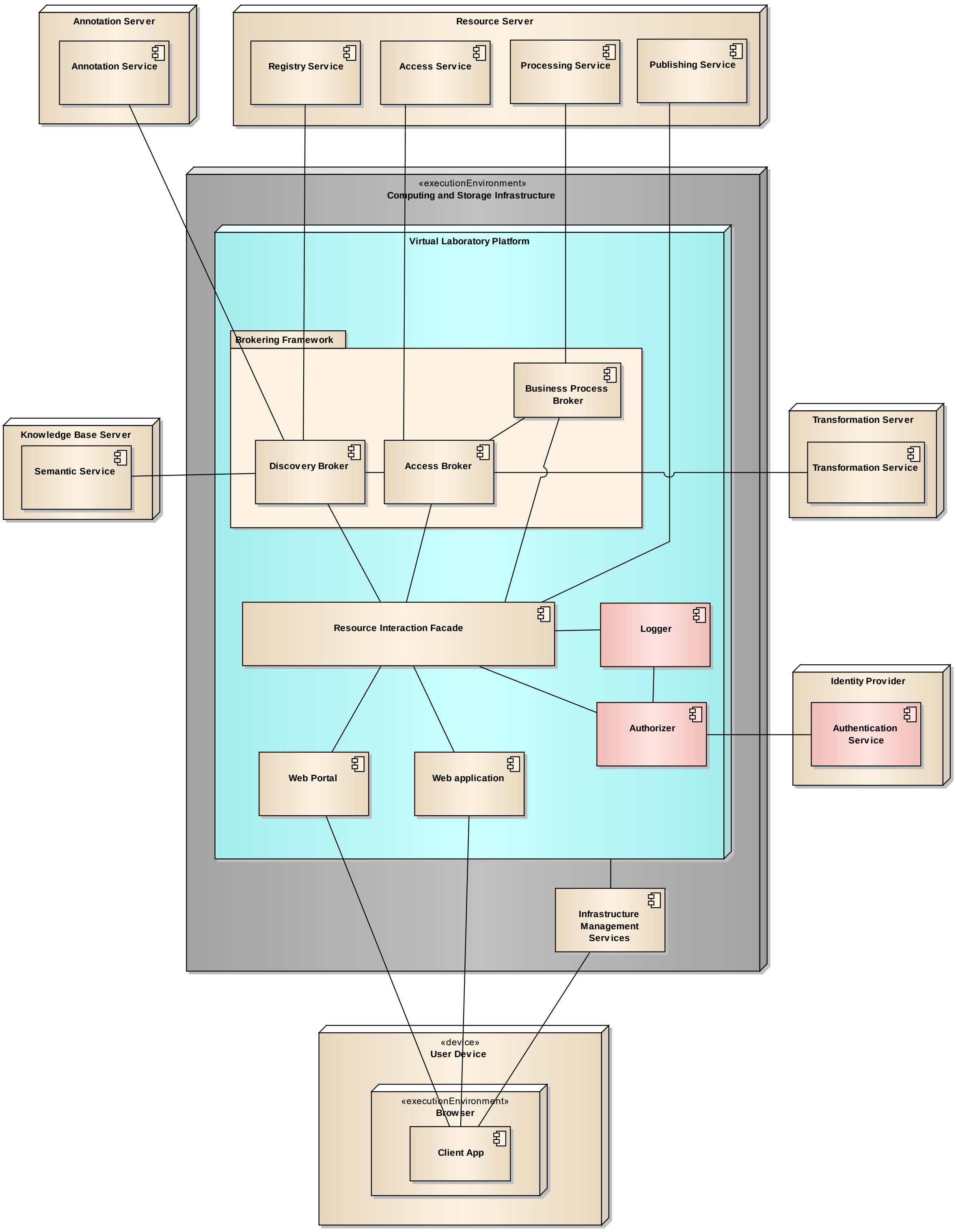

Figure 11 Engineering view of the ECOPOTENTIAL Virtual Laboratory architecture

The ECOPOTENTIAL architecture includes a set of different nodes:

Node	Description
Resource Server	A Resource Server is a node dedicated to serve resources. ECOPOTENTIAL assumes that Resource Server nodes are existing, up and accessible, and providing at least the Registry Service. According to the brokering approach, no assumption is made about communication protocols.

Annotation Server	An Annotation Server is a node hosting an Annotation Service. It may be deployed either inside or outside of the Computing and Storage Infrastructure.
Virtual Laboratory Platform	A Virtual Laboratory is the core architectural node. It contains all the components and tools needed to achieve the ECOPOTENTIAL objective.
Computing and Storage Infrastructure	A Computing and Storage Infrastructure is a node hosting the Virtual Laboratory Platform. It may be either a node managed locally by one ECOPOTENTIAL partner, or a private or public cloud offering Infrastructure-as-a-Service capabilities.
Knowledge Base Server	A Knowledge Base Server is a node providing services accessible by the Virtual Laboratory for semantic enhancements. According to the brokering approach, no assumption is made about communication protocols.
Transformation Server	A Transformation Server is a node providing services accessible by the Virtual Laboratory for data transformations (e.g. re-projection on different Coordinate Reference Systems, format encoding, sub-setting, change of resolution and interpolation) enhancements. According to the brokering approach, no assumption is made about communication protocols.
User Device	A User Device is a node hosting user's applications. It can be a desktop, or a mobile device. The only assumption is that it is able to host a (modern) Web browser.
Identity Provider	An Identity Provider is a node hosting an Authentication Service

Such nodes collectively host the software components interacting for an easier use of ecosystem resources:

Component	Description
Annotation Service	Annotation Services enable publishing and visualization of annotation with a proper interface supported by the Discovery Broker
Registry Service	Registry Services enable the discovery of resources. They range from simple inventories listing the available resources, to full catalogue services processing complex queries.
Access Service	Access Services enable the download of resource representations (e.g. datasets, workflow diagrams, etc.). They range from plain download services to full access services supporting subsetting, interpolation, re-projection, format transformations. Access Services include provision of graphical representations of resources, i.e. visualization services.
Publishing Service	Publishing Services enable the upload of resources (e.g. datasets, workflow diagrams, etc.).
Processing Service	Processing Services enable the execution of processing algorithms and workflows, generating products that are stored as new resources.
Brokering Framework	At the core of the Virtual Laboratory the Brokering Framework package includes a set of components which harmonize discovery and access of heterogeneous resources. It includes at least: • A Discovery Broker which connect with many different discovery, registry and inventory services, exposing several standard or well-known discovery interfaces. Through this well-known interfaces, a user can discovery all the datasets published by the different data sources. • Support for semantic enhancement of discovery. A simple query can be expanded in multiple queries based on the semantics relationship defined in an external knowledge base. • An Access Broker which connects with many different access and download services, exposing several standard or well-known access interfaces. Through these well-known interfaces, a user can access all the datasets published by the different data sources. • Support for data transformation. Multiple datasets can be transformed accessing external transformation services, in order to harmonize them on the same Common Grid Environment (same

	spatial and temporal coverage, same resolution, same Coordinate Reference System, same data format, etc.) • A Business Process Broker which gets a BPMN representation of a scientific business process - a ECOPOTENTIAL Workflow – translates it in an executable process (e.g. expressed in BPEL) and runs it.
Semantic Service	Semantic Services expose knowledge-bases such as thesauri, gazetteers, ontologies, allowing to find terms related to a keyword for query expansion.
Transformation Service	Transformation Services implement datasets transformation (e.g. subsetting, re-projection, interpolation)
Resource Interaction Facade	The Resource Interaction Facade is a component that aggregates the different components exposing a simplified interface facilitating the interaction with the VLab Platform
Infrastructure Management Services	The Infrastructure Management Services are provided by the Computing and Storage Infrastructure. They allow a VLab Administrator to manage the VLab Platform instance through a browser.
Web Portal	A Web Portal is the primary interface for Human-to-Machine interaction. It allows at least discovery, upload and download of datasets for offline usage.
Web Application	A Web Application is a specific component implementing (part of) the application logic of a Web or mobile app. It implements the needed workflow interacting with the Brokering Framework, the Data Publisher, etc.
Browser	The Browser is the component enabling user's interaction with the system. It will host part of the application logic (as client-side code) and the presentation logic.

The following table lists the main security components:

Component	Description
Authentication Service	The Authentication Service, hosted in the Identity Provider node, verifies user's identity. It is contacted by the Web portal or applications for sending credentials, and it can be contacted by the Authorizer for verification
Authorizer	The Authorizer is a software component receiving requests from the Web portal or applications and making decisions about allowing/denying actions.
Logger	The Logger reports the actions requested to the VLab Platform

5.6 Technology Viewpoint

Technology VP is concerned with the choice of technology to support system distribution.
[28]

The ECOPOTENTIAL system will be implemented using and extending existing solutions and tools. At the time of the proposal and DoW preparation some key technologies were preliminarily identified. They were mostly provided or under control of ECOPOTENTIAL partners. Other potential technologies have been identified during the preparation of the ECOPOTENTIAL DMP.

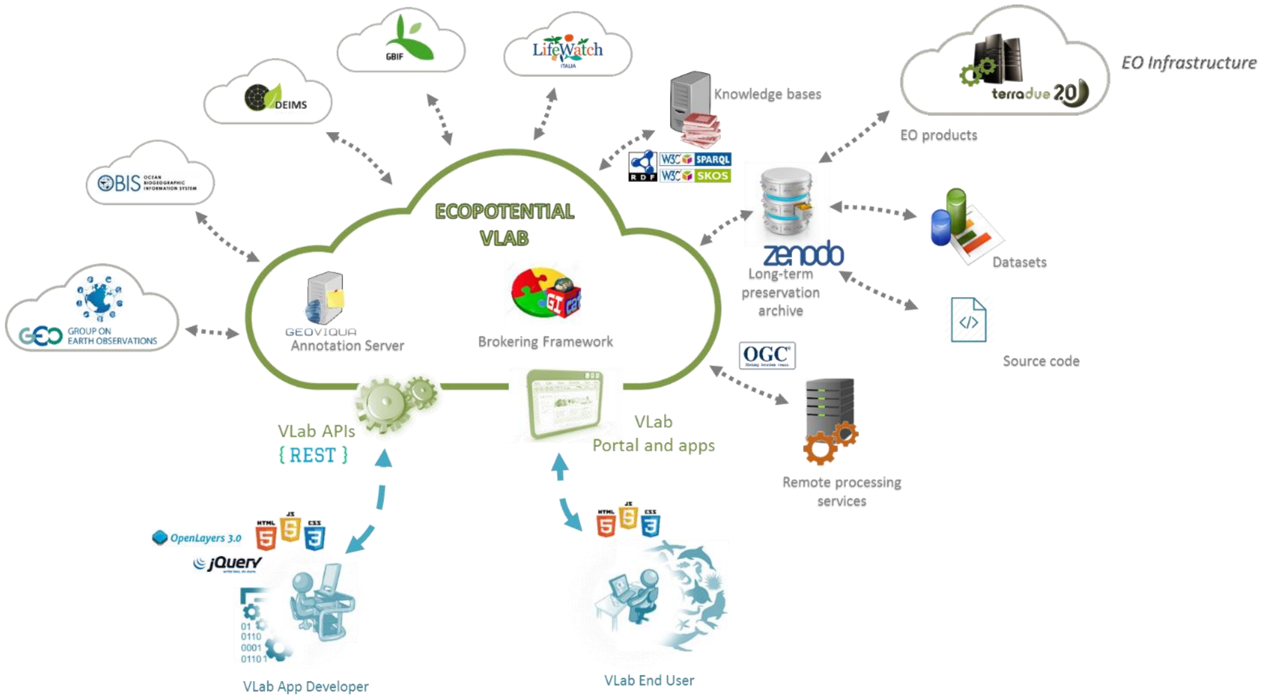

Figure 12 Virtual Laboratory technologies (not exhaustive)

5.6.1 Brokering Framework

The main technological choice regarding the architecture and implementation of the Virtual Laboratory is the selection of the brokering framework. The proposal and the DoW identified the GI-Suite Brokering Framework as the best solution for maturity and interoperability with GEOSS.

The **GI-suite Brokering Framework** is a suite of technologies developed by CNR-IIA to implement an information Brokering Framework that allows for uniform semantically enriched discovery and access to heterogeneous geospatial data sources; multidisciplinary interoperability integrating GIS and EO data from multiple infrastructures (e.g. INSPIRE compliant, Copernicus services).

The suite is composed of the following components:

- GI-cat: a discovery broker;
- GI-sem: a semantic broker;
- GI-axe: an access broker;
- GI-quality: a quality broker;
- GI-BP: a business process broker.
- GI-portal: a Web (thin) client to test the suite;
- GI-APIs: high-level JavaScript APIs to make use of the brokering suite.

The GI-suite Brokering Framework supports access through several interfaces including: OGC WCS (1.0.0, 1.1.2 & 2.0.1), OGC WMS (1.1.1, 1.3.0), OGC WFS (1.0.0, 1.1.0), FTP, WAF, NetCDF CF (1.6), HDF, CUAHSI HIS Server, THREDDS (1.0.1, 1.0.2), OPeNDAP, File system, Environment Canada Real-time Hydrometric Data FTP and BCODMO. It supports queries through several interfaces including: OGC WCS (1.0.0, 2.0.1), OGC WMS (1.1.1, 1.3.0), OGC WFS (1.0.0), OGC WPS (1.0.0), OGC SOS (1.0.0), CUAHSI HIS Server, ArcGIS REST API. (See section 6.2.1 for a full list of supported protocols.))

The GI-suite Brokering Framework has already being used in several projects and has been improved through them (EuroGEOSS, GEOWOW, ENVIROFI, GeoViQua, ODIP). The EuroGEOSS Brokering Framework was actually the basis of the current GI-suite Brokering Framework where the concept of query expansion enabled in the Brokering Framework accessing semantic assets (vocabularies, thesauri, ontologies) stored in a knowledge base was introduced. In the ENVIROFI project, access-brokering capabilities were enhanced and in the GeoViQua it was extended to integrate quality information provided by data producers, and feedback from users. It is mature enough and extensible, allowing for the integration of new capacities needed by the ECOPOTENTIAL project as identified in WP4. It has been adopted in operational settings like GEOSS. (See Section 6.2.1 for a detailed description of the GI-suite modules used in ECOPOTENTIAL.)

The choice of the GI-suite Brokering Framework as the central brokering component of the VLabs is determined firstly by its features. In particular, it is specifically designed to integrate geospatial services from heterogeneous domains like those cited in the call (INSPIRE, Copernicus, etc.). Secondly, its maturity has been proven by its use in several European Projects and by its adoption in initiatives such as GEOSS (with the development of the GEO-DAB). Its functionality has been increasing since the moment of suggesting its use in the ECOPOTENTIAL proposal. Thirdly and not less important, it is under continuous incremental development by one of the partners, CNR-IIA, so the control to include the new functionalities needed to cover the requirements established by WP3-WP9 within the ECOPOTENTIAL consortium.

5.6.2 Semantic Service

The GI-Suite Brokering Framework is able to connect to external semantic services. It currently supports SKOS (Simple Knowledge Organization System) knowledge bases publishing a SPARQL (SPARQL Protocol and RDF Query Language) interface. It is tested with the EC-JRC semantic service adopted in GEOSS.

EC-JRC Semantic service. The SemanticLab of the Institute for Environment and Sustainability (ISE) of the European Commission Joint Research Center (EC-JRC) developed a semantic service providing access through a SPARQL (SPARQL Protocol and RDF Query Language) interface to a knowledge base structured according to SKOS (Simple Knowledge Organization System) and encoded in RDF (Resource Description Framework). The knowledge base includes a set of aligned thesauri and ontologies:

"AIP-3-Hydrosphere Vocabulary, version 1.0"@en	http://www.cuahsi.org/navigation/hydrosphere
"CaLAThe-Cadastre and Land Administration Thesaurus, version 1.0"@en	http://www.cadastralvocabulary.org/CaLAThe
"EUROVOC v4.3"@en	http://eurovoc.europa.eu/EUROVOC/v4.3
"EuroGEOSS-Drought Vocabulary, version 1.0"@en	http://eurogeoss.eu/DroughtVocabulary
"GCMD-Earth Science Keywords, version 5.3.3"@en	http://gcmd.gsfc.nasa.gov/skos
"GCMD-Earth Science Keywords, version 5.3.3"@en	http://gcmd.gsfc.nasa.gov/skos
"GEMET-INSPIRE themes, version 1.0"@en	http://inspire-registry.jrc.ec.europa.eu/registers/Themes/items
"GEMET, version 2.4"@en	http://www.eionet.europa.eu/gemet/
"GEOSS - Earth Observation Vocabulary, version 1.0"@en	http://www.earthobservations.org/GEOSS/EO_Vocabulary
"GEOSS - Societal Benefit Areas, version 1.0"@en	http://iaaa.unizar.es/thesaurus/SBA_EuroGEOSS
"INSPIRE-Feature Concept Dictionary, version 3"@en	http://inspire-registry.jrc.ec.europa.eu/registers/FCD/items
"INSPIRE-Glossary, version 3"@en	http://inspire-registry.jrc.ec.europa.eu/registers/GLOSSARY/items
"ISO-19119 geographic services taxonomy"@en	http://inspire-registry.jrc.ec.europa.eu/registers/ISO_19119/items

The semantic service published by EC-JRC and providing a set of aligned thesauri will be initially used for multilingualism, suggestions, and semantic queries.

Whenever required, other knowledge base can be developed and published using open source tools supporting SPARQL/SKOS.

5.6.3 Transformation Service

The GI-suite Brokering Framework already supports subsetting and simple interpolation schemes, and the most used CRS through either local routines or external transformation services.

5.6.4 Resource Interaction Facade

The Resource Interaction Facade is an ancillary component aiming to expose a harmonized and consistent interface to the many components of the ECOPOTENTIAL architecture – e.g. the Brokering Framework, the Processing Service(s) and Publishing Service(s) - according to the Facade pattern. The component will be developed in ECOPOTENTIAL, and the Web Portal and Web Application(s) will use the APIs exposed by the Resource Interaction Façade.

5.6.5 Web Portal, Web Applications and Client Apps

As usual in the modern Web environment, different components realize the user applications: server-side components (Web Portal and Web Applications) running on the VLab platform and client components (Client Apps) running in the user's browser. This allows splitting the business logic between server and clients for achieving better performances. The Web Portal and the Web Applications carry out part of the business logic and deliver both presentation content – e.g. HTML and CSS code – and mobile code – i.e. Javascript code – to the browser. The interaction between server-side and client-side components realize the application. In this way, the processing load and responsibility can be freely allocated between server and client, allowing a wide range of options: from full server applications where client only manages the user interface – e.g. no Javascript code – to full client applications, like typical Web and mobile apps, where the server only dispatches call to internal components.

To support such a programming model the VLab Platform exposes a clean server-side API through the Resource Interaction Facade. The server-side API bases on REST architectural style, and JSON encoding. A simple Javascript library including HTML+CSS widgets facilitate client-side interaction. Developers can create server-side components communicating with the VLab through the server-side API, and client-side components using the Javascript library.

5.6.6 Resource Server(s)

As the name implies, Resource Servers provide access to resources. In particular, they include one or more of the following internal components:

- Registry Service
- Access Service
- Publishing Service
- Processing Service

The ECOPOTENTIAL architectural approach, based on brokering System of Systems, poses no constraint about the communication protocols with the different services. The technological choices are then guided by the characteristics of the different resources managed in ECOPOTENTIAL.

Satellite data and derived products

For satellite data the main requirement was the long-term preservation capabilities. ECOPOTENTIAL selected Zennodo as the main sharing platform for satellite data and derived products. Zenodo is an open dependable home for the long-tail of science, enabling researchers to share and preserve any research outputs in any size, any format and from any science. Zenodo content is accessible through both REST API and OAI-PMH API providing Registry and Access Services.

In-situ data

For in-situ, it is expected that many data will be available through existing data systems including: LifeWatch (E-Science European Infrastructure for Biodiversity and Ecosystem Research), OBIS (Ocean Biogeographic Information System) and GBIF (Global Biodiversity Information Facility). Datasets generated in ECOPOTENTIAL will be made available through Zenodo (netCDF files) or OGC compliant systems.

Model results

Model results are another category of datasets. They will be published on Zenodo or other data system for satellite or in-situ data (see above).

Indicators/indices

Model results are another category of datasets. They will be published on Zenodo or other data system for satellite or in-situ data (see above).

Processing algorithms

The stable versions of processing algorithms generated in the project and encoded in any programming language will be published in either source or executable version on Zenodo. As a long-term preservation platform, Zenodo assures that users can always refer to a specific version of the algorithm. Unstable versions and beta versions may be stored on GitHub supporting collaborative development.

Models/workflows descriptions

The stable versions of models/workflows descriptions in BPMN will be published on Zenodo. As a long-term preservation platform, Zenodo assures that users can always refer to a specific version of the model/workflow.

Processing services

Processing services are exposed through documented protocols (e.g. OGC WPS profiles). ECOPOTENTIAL provides access at least at the Terradue platform for EO data processing.

The GI-Suite Brokering Framework is already able to connect the majority of existing data systems, repositories and services cited above. For Zenodo a specific accessor will be developed.

5.6.7 Annotation Server

The Annotation Server will be realized extending the component implemented in the European project ConnectinGEO.

6 IMPLEMENTATION

6.1 Development approach

In early 2000, new software design and development methodologies were proposed, with the objective of solving issues emerged in traditional software engineering approaches such as the waterfall model (Figure 13) [32] and other sequential processes, in particular with the advent of the Internet and related Web applications. Those new development methodologies shared a set of principles defined in the Manifesto for Agile Software Development (Agile Manifesto) [33]:

- Individuals and interactions over processes and tools

- Working software over comprehensive documentation
- Customer collaboration over contract negotiation
- Responding to change over following a plan

As an innovation project aiming at facilitating the use of data by users in a highly dynamic and evolving sector, ECOPOTENTIAL has great requirements at least on privileging "working software", "customer collaboration" and fast "response to change". Therefore, ECOPOTENTIAL will adopt an Agile Methodology for design and development.

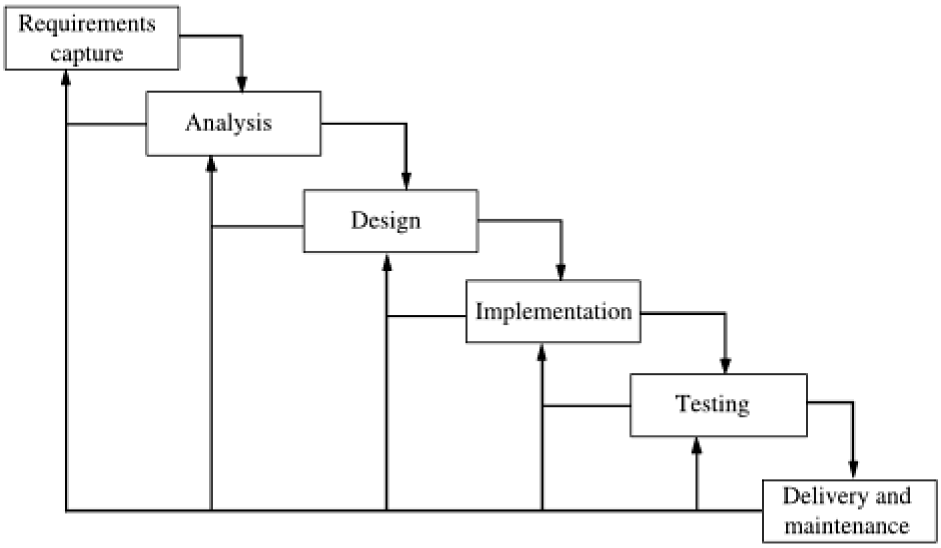

Figure 13 The traditional Waterfall Model (from [32])

Agile methodologies better respond to changes through an iterative process (Figure 14). Requirements are not entirely collected at the beginning of the process as in the traditional processes. They may be added later to be fulfilled in a next iteration.

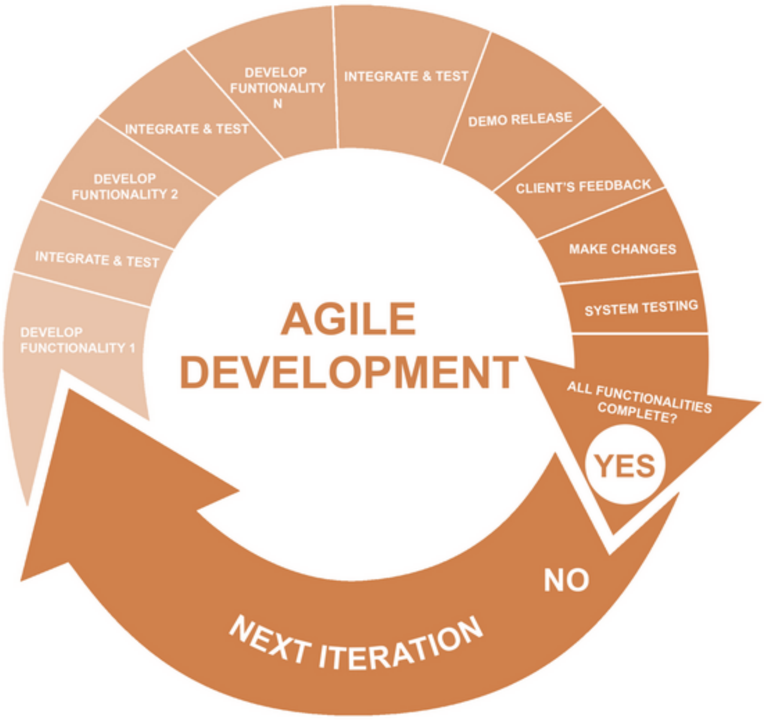

Figure 14 The iterative process in Agile development

Taking into account the specificity of ECOPOTENTIAL we can identify four main milestones and therefore four main iterations (see Figure 15):

- Project Month 12, end of the first major iteration and release of the Metadata Platform. The Metadata Platform provides metadata harmonization and data access from heterogeneous data sources.
- Project Month 24, end of the second major iteration and release of the Workflow Platform. The Workflow Platform provides the possibility to run ECOPOTENTIAL workflows on selected datasets.
- Project Month 36, end of the third major iteration and release of the User-focused Services Platform. The User-focused Services Platform provides all the ECOPOTENTIAL services allowing the implementation of user-driven portal and apps.
- Project Month 48, end of the fourth major iteration and release of the final ECOPOTENTIAL VLab Platform. The ECOPOTENTIAL VLab Platform provides the same capabilities of the User-focused Services Platform but with improved performances after test and evaluation.

Each iteration includes the following phases:

1) Definition and prioritization of functionalities based on collected requirements and feedback
2) Cycle over the selected functionalities for the iteration:
   a. Development of functionality
   b. Integration and test
3) Demo release
4) Collection of feedback from the consortium and presentations in external events
5) Release of the VLab capacity

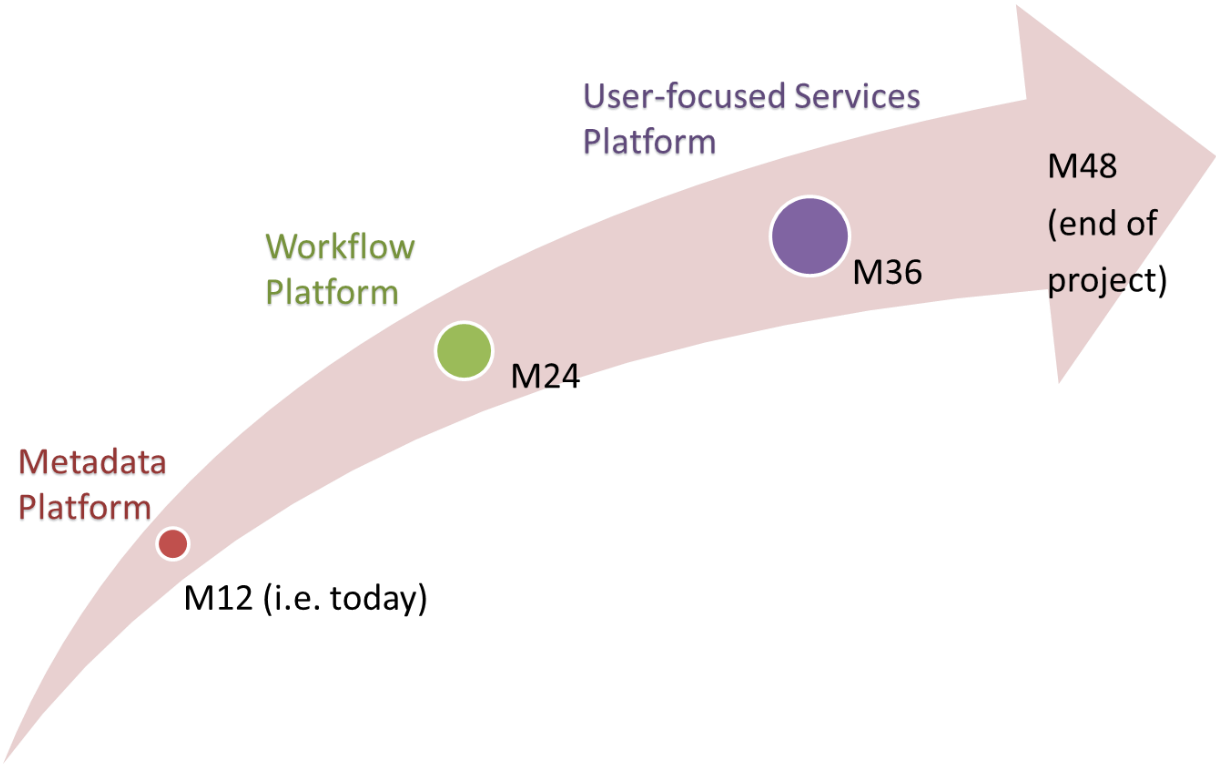

Figure 15 Main iterations of the ECOPOTENTIAL VLab implementation

6.2 System integration

As described in Section 4.1, ECOPOTENTIAL Virtual Laboratory adopt an Open Architecture with Decentralized Software Evolution based on APIs allowing internal integration of existing tools and external interaction with other members of the geospatial ecosystem. The different components of the Virtual Laboratory architecture are then implemented through the integration of selected technological solutions to build a complete framework delivering the requested Virtual Laboratory functionalities. The first release of the Virtual Laboratory comprised the GI-suite Brokering Framework. The following releases will include selected components integrated with the GI-suite Brokering Framework to support missing functionalities. In particular, during the second and third iterations the following components will be integrated: a) Annotation Server from the ConnectinGEO project b) Federated identity systems (e.g. Google, Facebook, other OpenAuth systems) for authentication.

6.2.1 The GI-suite Brokering Framework

The GI-suite Brokering Framework is a set of coordinated software components for geospatial resource brokering. The main components used in ECOPOTENTIAL are:

- *Discovery broker (GI-cat)*: a component which is able to connect disparate (distributed and heterogeneous) metadata sources, exposing them through a set of standard catalogue interfaces. By means of metadata harmonization and protocol adaptation, it is able to search metadata from different sources and transform query results to a uniform and consistent metadata model. GI-cat mediates among the connected metadata sources interfaces, and harmonizes their metadata mapping them to an internal schema based on ISO 19115 (GI-cat metadata model). Each query request sent through the external interfaces is performed against all the connected sources based on the internal schema. GI-cat supports both distributed queries (for external sources exposing a catalogue service) and harvesting. Harvesting can be adopted for enhance query performances for catalogues, or to enable search also on inventory services providing metadata without catalogue functionalities. The choice between distributed query and harvesting can be made per data sources. In case of harvesting also the repetition time can be defined per data source. Internally, GI-cat includes several modules (see Figure 16):

- The *Distributor* is in charge of accepting queries from the exposed catalogue interfaces and route them to the external data sources. The Distributor accesses the Local DB for harvested data sources, and Accessors for query propagation.
- The *Profilers* are adaptors for exposing catalogue interfaces to users. Each Profiler exposes a standard interface carrying out: a) mapping of the query interface to the internal search interface of the Distributor; b) mapping of metadata from the GI-cat metadata model to the metadata model of the supported interface, providing also the related encoding. For example the CSW/ISO Profiler maps the OGC Catalog Service for Web (CSW) interface to the internal search interface, and, on the other direction, it maps the metadata from the internal model to the ISO 19115 model and ISO 19139 encoding.
- The *Harvesters* periodically harvest the related data source filling the Local DB.
- The *Accessors* are adaptors for connecting metadata sources. Each Accessor supports a metadata source carrying out: a) mapping of an internal query (from query propagation or harvesting) to the interface exposed by the external metadata source; b) mapping of resulting metadata to the GI-cat metadata model. For example, the Accessor for Web Accessible Folder WAF hosting ISO 19139 XML files, maps the request (only from harvesting since WAF is an inventory service and not a catalogue service) to a HTTP request, and on the other direction, it maps the metadata from ISO 19139 (ISO 19115 model) to the GI-cat metadata model.
- The *Local DB* hosts the harvested metadata.

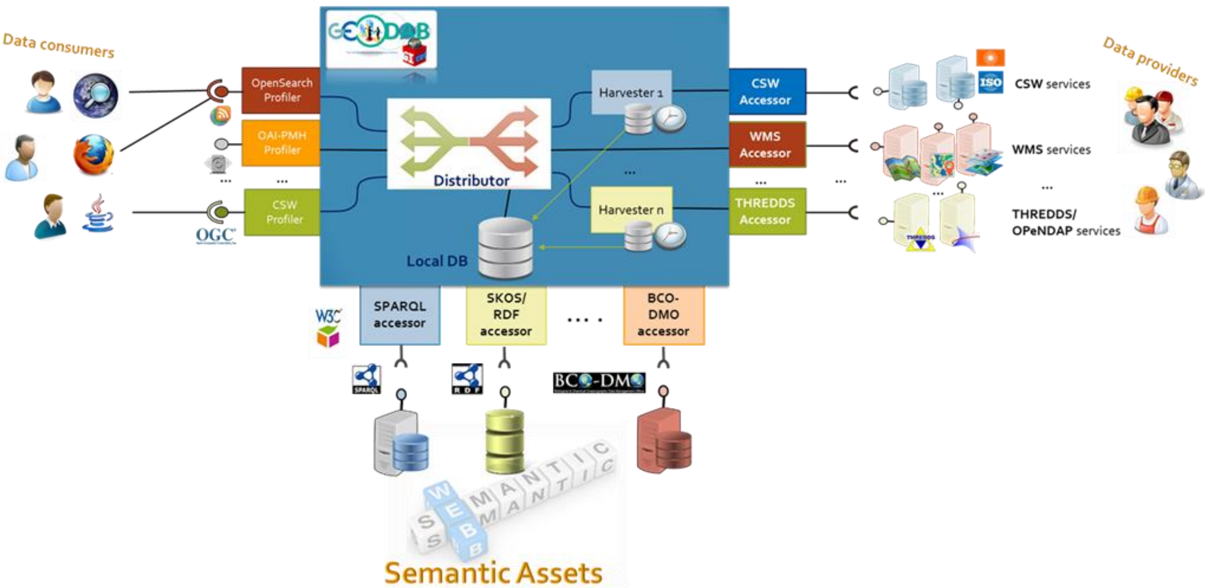

Figure 16 Discovery broker (GI-cat) internal architecture

- *Semantic Enhancement Module (GI-sem)*: a component which implements semantic query expansion [31]. If the semantic query is enabled by configuration, when a query includes a keyword, it is passed as a parameter of a semantic query to a set of connected knowledge bases to search for "related" terms. Each of the resulting term is then used as a keyword in a separate geospatial query. The results are then assembled to provide the complete response to the user. This workflow enables several semantic enhancements depending on the connected knowledge bases, including multilingualism, semantic refinements and suggestions. For example, connecting a multilingual thesaurus supporting English, French, German, Italian, Polish and Spanish, if an user send a request for "moisture" in English, then several separate geospatial queries will be sent through GI-cat, including for "moisture" (English), "humidité" (French), "Feuchtigkeit" (German), "umidità" (Italian), "wilgoć" (Polish) and "humedad" (Spanish). This allows to find datasets annotated in different languages overcoming limitations of syntactic queries on metadata content. GI-sem supports basic relationships such as "related" (i.e. generic relationship; e.g. "soil moisture" is related to "soaking"), "broader" (i.e. generalization; e.g. "soil water" is more general than "soil moisture") or "narrower" (i.e. specification; e.g. "soil moisture" is more specific than "soil water"). GI-sem is implemented through semantic accessors integrated in GI-cat, which map the request to a specific knowledge base interface.
- *Access broker (GI-axe)*: a component which is able to connect with disparate (distributed and heterogeneous) data sources, exposing them through a set of standard catalogue interfaces. By means of data harmonization and protocol adaptation, it is able to download (subset of) datasets from different sources. GI-axe mediates among the connected data sources interfaces, and harmonizes datasets using a small set of internal data models (GI-axe data models). It is also able to carry out on-the-fly transformations for subsetting, reprojection, resampling, encoding. Internally, GI-cat includes several modules (see Figure 17):
  - The *Orchestrator* is in charge of accepting data access requests from the exposed data access interfaces and run the needed workflow for access and transformation. The Orchestrator is a smart component taking into account servers' capabilities: if the original data source already supports the requested transformation, the Orchestrator relies on it, otherwise it calls the Converters.
  - The *Profilers* are adaptors for exposing access interfaces to users. Each Profiler exposes a standard interface carrying out: a) mapping of the data access interface to the internal access interface of the Orchestrator; b) mapping of datasets from the GI-axe data models to the data model of the supported interface, providing also the related encoding. For example the WCS/netCDF Profiler maps the OGC Web Coverage Service (WCS) interface to the internal access interface, and, on the other direction, it transforms the dataset from the GI-axe data model to the netCDF data model and encoding.
  - The *Accessors* are adaptors for connecting data sources. Each Accessor supports a data source carrying out: a) mapping of an internal access request to the interface exposed by the external data source; b) mapping of resulting datasets to the GI-axe data model. For example, the Accessor for FTP hosting GeoTIFF files, maps the data access request to a FTP download request, and on the other direction, it transforms the GeoTIFF dataset to the GI-axe data model.
  - The *Converters* are modules for on-the-fly execution of dataset transformations. These transformations include simple processing aiming not to modify the content of datasets, but only to transform its representation. They include subsetting, reprojection, resampling and encoding. The Converters either use local routines or call external web services exposed through OGC Web Processing Service (WPS) interface.

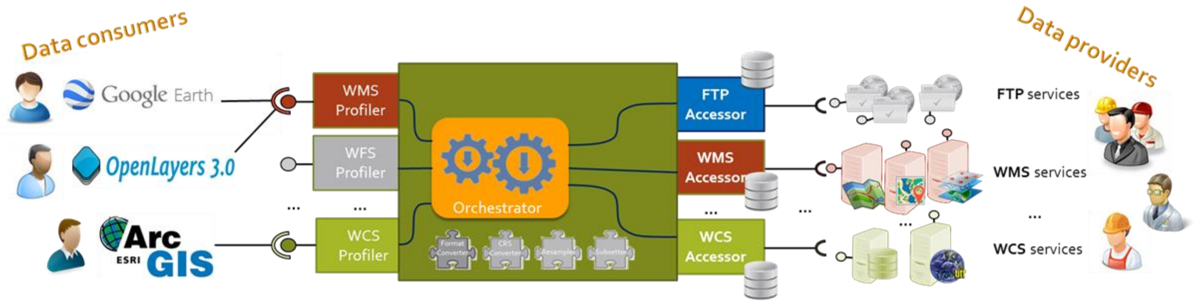

Figure 17 Access Broker (GI-axe) internal architecture

- *Configurator (GI-conf)*: a user friendly web tool which allows the Brokering Framework configuration using a browser. With GI-conf an administrator can manage the published interfaces, the brokered sources and edit several other settings such as proxy parameters and personalize the welcome page.

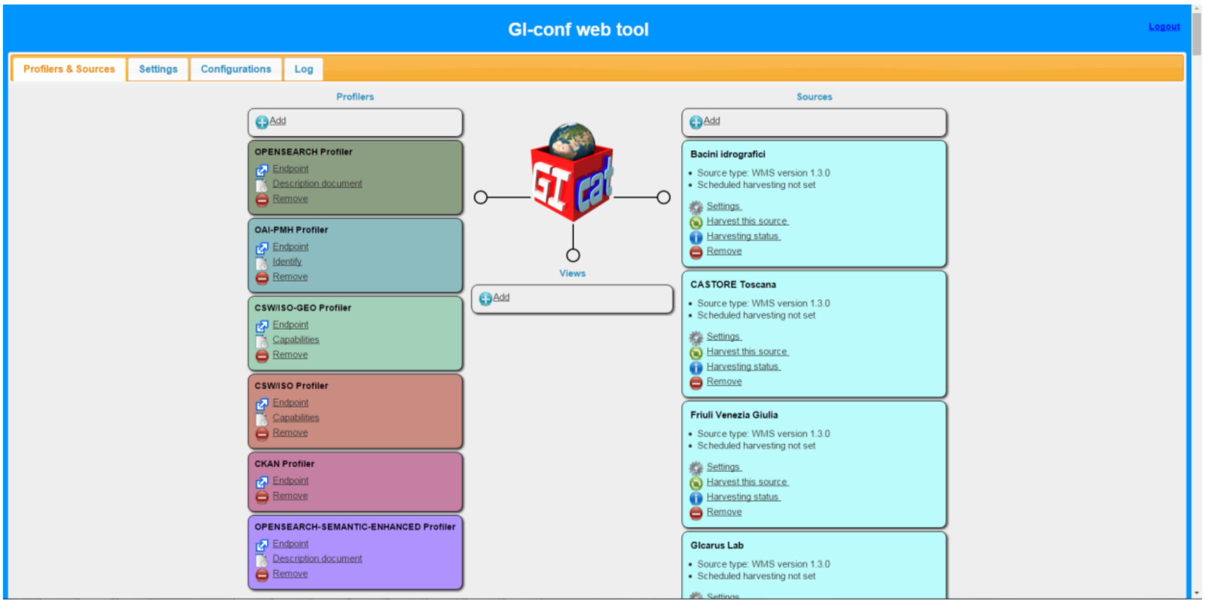

Figure 18 GI-conf screenshot

- *Business Process Broker*: a component which is able to analyze a BPMN representation of an abstract business process, to compile it in an executable BPEL instance - adding components as necessary - and run it.
- *Test Portal (GI-Portal)*: a basic portal for testing the GI-suite Brokering Framework capabilities, operation and configuration.
- *Application Programming Interface (GI-API)*: a Javascript library implementing Web APIs for interaction with the GI-suite Brokering Framework. It is conceived as a set of objects and related methods to simply use the Brokering Framework capabilities for rapid development of Web and mobile applications (documentation available at http://api.eurogeoss-broker.eu/docs/index.html).

Table 8 shows the data sources (accessors for discovery and access) currently supported by the GI-suite Brokering Framework.

Protocol	Protocol elements
OGC WCS OGC WCS 1.0, 1.1, 1.1.2	Discovery (coverages inventory) and access interfaces
OGC WMS OGC WMS 1.3.0, 1.1.1	Discovery (maps inventory) and access interfaces
OGC WFS OGC WFS 1.0.0	Discovery (features inventory) and access interfaces
OGC WPS OGC WPS 1.0.0	Discovery (processes inventory) and access interfaces
OGC SOS 1.0.0	Discovery (sensors inventory) and access interfaces
OGC CSW CORE OGC CSW 2.0.2 Core, OGC CSW ISO AP ISO 1.0, OGC CSW ebRIM CIM ebRIM/CIM, OGC CSW ebRIM EO ebRIM/EO, CWIC	Discovery interface and metadata profiles
fr FLICKR	Discovery and access interfaces
HDF HDF	Metadata and data encoding
HMA HMA CSW 2.0.2 ebRIM/CIM	Discovery interface
GeoNetwork (versions 2.2.0 and 2.4.1) catalog service	Discovery interface
dee gree Deegree (version 2.2) catalog service	Discovery interface
ESRI ArcGIS Geoportal (version 10) catalog service	Discovery interface
WAF Web Accessible Folders 1.0	Discovery and access interfaces and metadata model
FTP - File Transfer Protocol services populated with supported metadata	Discovery and access interfaces
THREDDS 1.0.1, 1.0.2	Discovery and access interfaces
ncISO THREDDS-NCISO 1.0.1, 1.0.2	Discovery and access interfaces, and metadata model
ncISO THREDDS-NCISO-PLUS 1.0.1, 1.0.2	Discovery and access interfaces, and metadata model
CDI 1.04, 1.3, 1.4 1.6	Discovery interface and metadata model
GI-cat 6.x, 7.x	Discovery and access interfaces
GBIF	Discovery and access interfaces, and metadata model
OpenSearch 1.1 accessor	Discovery interface
OAI-PMH 2.0 (support to ISO19139 and dublin core formats)	Discovery interface and metadata model
NetCDF-CF 1.4	Metadata and data model
NCML-CF	Metadata and data model

NCML-OD	Metadata and data model
ISO19115-2	Metadata model
GeoRSS 2.0	Access interface, and metadata model
GDACS	Access interface, metadata and data models
DIF	Metadata and data model
File system	Access interface
SITAD (Sistema Informativo Territoriale Ambientale Diffuso) accessor	Discovery and access interfaces
INPE	Discovery and access interfaces
HYDRO	Discovery and access interfaces
EGASKRO	Discovery and access interfaces
RASAQM	Discovery and access interfaces
IRIS event	Discovery and access interfaces, metadata model
IRIS station	Discovery and access interfaces, metadata model
UNAVCO	Discovery and access interfaces, metadata model
KISTERS Web - Environment of Canada	Discovery and access interfaces
DCAT	Discovery interface and metadata model
CKAN	Discovery interface and metadata model
HYRAX THREDDS SERVER 1.9	Discovery and access interfaces

Table 8 Preliminary list of data sources protocols supported by the GI-Suite Brokering Framework

Table 9 shows the protocols for the exposed interfaces (discovery and access profilers) currently supported by the GI-suite Brokering Framework.

Protocol	Protocol elements
OGC CSW 2.0.2 AP ISO 1.0	Discovery interface and metadata
OGC CSW 2.0.2 ebRIM EO	Discovery interface and metadata
OGC CSW 2.0.2 ebRIM CIM	Discovery interface and metadata
ESRI GEOPORTAL 10	Discovery and access interfaces
OAI-PMH 2.0	Discovery and access interfaces

OpenSearch 1.1 (including mapping to Atom)	Discovery interface and metadata model
OpenSearch 1.1 ESIP (including mapping to Atom)	Discovery interface and metadata model
OpenSearch GENESI DR	Discovery interface
GI-cat extended interface	Discovery and access interfaces
CKAN	Discovery and access interfaces, metadata model

Table 9 Preliminary list of protocols for published interfaces supported by the GI-Suite Brokering Framework

The GI-suite Brokering Framework is developed in Java language (for server-side components) and HTML+CSS+Javascript (for client-side components) and it is available in Web ARchive Format (WAR) for deployment in Java Servlet containers, such as Apache Tomcat and Jetty. It is currently adopted in several contexts (see Table 10), with different deployment strategies including local infrastructures with web application servers based on different servlet containers, private clouds adopting different virtualization techniques, public commercial cloud providing Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS) capabilities like Amazon.

EU BON	EU-BON Homepage	EU BON - Building the European Biodiversity Observation Network. EU BON proposes an innovative approach in terms of integration of biodiversity information system from on-ground to remote sensing data, for addressing policy and information needs in a timely and customized way. GI-cat is used as the EU-BON metadata registry.
CEOS Water Portal	CEOS Water Portal	CEOS Water Portal led by Japan Aerospace Exploration Agency (JAXA) is a project of the Applications Subgroup of the Committee on Earth Observation Satellites (CEOS) Working Group on Information Systems and Services (WGISS). The purpose of the CEOS Water Portal Project is to provide assistance to the water relevant scientists and general users (or non-researchers) in the development of data services associated with data integration and distribution.
GMOS	GMOS	The Global Mercury Observation System (GMOS) is aimed to establish a worldwide observation system for the measurement of atmospheric mercury in ambient air and precipitation samples. GMOS will include ground-based monitoring stations, shipboard measurements over the Pacific and Atlantic Oceans and European Seas, as well as aircraft-based measurements in the UTLS.
TREES 4 FUTURE	Trees 4 future	Trees4Future is an Integrative European Research Infrastructure project that aims to integrate, develop and improve major forest genetics and forestry research infrastructures. It will provide the wider European forestry research community with easy and comprehensive access to currently scattered sources of information (including genetic databanks, forest modelling tools and wood technology labs) and expertise.

	Pangaea	The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.
	NSIDC Acadis	The Advanced Cooperative Arctic Data & Information Service (ACADIS) manages data and is the gateway for all relevant Arctic physical, life, and social science data for the National Science Foundation (NSF) Division of Polar Programs (PLR) Arctic Research Program (ARC) research community.
	SeaDataNet FP6 project and SeaDataNet2 FP7 project	SeaDataNet objective is to construct a standardized system for managing the large and diverse data sets collected by the oceanographic fleets and the new automatic observation systems. The aim is to network and enhance the currently existing infrastructures, which are the national oceanographic data centres and satellite data centres of European riparian countries, active in data collection. The networking of these professional data centres, in a unique virtual data management system will provide integrated data sets of standardized quality on-line. SeaDataNet CSW interface
	GEOSS (GEO-DAB)	The Group on Earth Observations, GEO, was established by a series of three ministerial-level summits. It currently includes 68 member countries, the European Commission, and 46 participating organizations. The vision of GEO is to create a Global Earth Observation System of Systems (GEOSS) to help realize a future wherein decisions and actions for the benefit of humankind are informed via coordinated, comprehensive and sustained Earth observations and information. The Global Earth Observation System of Systems will provide decision-support tools to a wide variety of users. As with the Internet, GEOSS will be a global and flexible network of content providers allowing decision makers to access an extraordinary range of information at their desk. The IP3 was conceived as a way to exercise the process that has been defined for reaching interoperability arrangements. The 2nd Phase of the AIP will augment the GEOSS Initial Operating Capability previously established.
	GENESI-DR	GENESI-DR, (Ground European Network for Earth Science Interoperations - Digital Repositories), has the challenge of establishing open Earth Science Digital Repository access for European and world-wide science users. GENESI-DR shall operate, validate and optimise the integrated access and use available digital data repositories to demonstrate how Europe can best respond to the emerging global needs relating to the state of the Earth, a demand that is unsatisfied so far.
	GIIDA	GIIDA is a CNR initiative (inter-departmental project) aiming to the design and development a multidisciplinary infrastructure for the management, processing and evaluation of Earth and environmental data. GIIDA aim is to implement the Spatial Information Infrastructure (SII) of CNR for Environmental and Earth Observation data.GIIDA central catalog
	EuroGEOSS	EuroGEOSS demonstrates the added value to the scientific community and society of making existing geographic systems and applications interoperable and used within the GEOSS and INSPIRE frameworks. The project will build an initial operating capacity for a European Environment Earth Observation System in the three strategic areas of Drought, Forestry and biodiversity. The concept of inter-disciplinary interoperability requires research in advanced modelling from multi-scale heterogeneous data sources, expressing models as workflows of geo- processing

		components reusable by other communities, and ability to use natural language to interface with the models. EuroGEOSS portal
HMA HeterogeneousMissionsAccessibility	ESA HMA-T	The main objective of this ESA project is to involve the stakeholders, namely national space agencies, satellite or mission owners and operators, in a harmonization and standardization process of their ground segment services and related interfaces. HMA is the first project launched and overviewed by the GSCB.
AFROMAISON	AfroMaison	AFROMAISON aims to propose concrete strategies for integrated natural resources management in Africa in order to adapt to the consequences of climate change. AFROMAISON is funded by the 7th Framework Program of the European Union. It has a budget of 4 million euro and a runtime of 3 years (March 2011-2014). AfroMaison portal
ISPRA	http://www.isprambiente.gov.it/it	The Institute for Environmental Protection and Research, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale), has been established by Decree no. 112 of 25 June 2008, converted into Law no. 133 (with amendments) on 21 August 2008. ISPRA performs, with the inherent financial resources, equipment and personnel, the duties of: - ex-APAT, Italian Environment Protection and Technical Services Agency (article 38 of Legislative Decree no. 300, July 30, 1999, and subsequently amended); - ex-INFS, National Institute for Wildlife (Law no. 157 of February 11, 1992, and subsequently amended); - ex-ICRAM, Central Institute for Scientific and Technological Research applied to the Sea (Decree no. 496, article 1-bis, December 4, 1993, converted into Law no. 61, Article 1, January 21, 1994, with amendments). The Institute acts under the vigilance and policy guidance of the Italian Ministry for the Environment and the Protection of Land and Sea (Ministero dell'Ambiente e della Tutela del Territorio e del Mare).

Table 10 List of infrastructures and initiatives using the GI-suite Brokering Framework

The GI-suite Brokering Framework is extensible through an Accessor Development Kit (ADK) for the development of accessors.

The GI-suite Brokering Framework exposes server-side APIs for discovery and access through the Profilers. In particular, the GI-cat Profiler, providing functionalities beyond the usual discovery and access, including feedback for query monitoring, is suitable for integration in complex environment (such as a Virtual Laboratory). It also exposes APIs for configuration and notification. The GI APIs facilitate the use of discovery and access functionalities by intermediate users (developers).

6.2.2 Annotation Server

The Annotation Server is implemented using the GEO User Feedback system which has been developed as a part of the FP7 GeoViQua project and enhanced in the H2020 ConnectinGEO project [34]. It consists of:

- A user feedback server
- A feedback submission client
- A feedback search client

It aims to improve the ways in which geospatial experts can communicate user feedback about geospatial data registered in the GEOSS portal. In a federated environment, such as GEOSS, implementing feedback is challenging because existing approaches do not apply to federated resources. This deficiency is addressed by

GeoViQua's feedback model which can be implemented independently from the resources subject to user feedback.

6.3 Status of implementation

The status of implementation at project-month 13 (June 2016) is summarized in the following tables listing the major components:

Node	Status
Resource Server	Some resource servers (or platforms) have been identified: LifeWatch, OBIS, GBIF, DEIMS, GEOSS, Terradue EO Platform, Zenodo, ESA Scientific Hub, CNR ecosystem model server
Annotation Server	Not yet available
Virtual Laboratory Platform	First prototype available
Computing and Storage Infrastructure	The current Vlab prototype is available as an instance based on Apache Tomcat hosted on Amazon EC2/S3 IaaS
Knowledge Base Server	The current Vlab prototype accesses the KB service hosted by the EC-JRC
Transformation Server	No external transformation service available
User Device	Desktop, laptop, and tablet hosting a modern browser are supported
Identity Provider	No authentication implemented

Component	Status
Annotation Service	No annotation service implemented.
Registry Service	The DMP lists some registry services in particular for satellite data and products. Updates of the DMP needed for other resources.
Access Service	The DMP lists some registry services in particular for satellite data and products. Updates of the DMP needed for other resources.
Publishing Service	The current protoype supports Zenodo for the publication of multiple resource types
Processing Service	CNR ecosystem model services used for proofs-of-concepts. Terradue EO Platform targeted, but technical information needed for proofs-of-concepts
Brokering Framework	GI-Suite Brokering Framework integrated in the VLab. It currently supports most of the protocols described in the DMP for interacting with Resource Servers. Accessors developed for Zenodo and CNR ecosystem model services.
Semantic Service	The VLab currently access the knowledge service published by the EC-JRC
Transformation Service	The VLab currently uses the GI-Suite Brokering Framework internal routines for dataset transformation (subsetting, re-projection, interpolation, encoding)
Resource Interaction Facade	The Vlab currently use the GI-Suite Brokering Framework simple APIs. Enhancements needed to support resource publishing through the VLab (if necessary), and workflow invocation.
Infrastructure Management Services	The Infrastructure Management Services are currently provided by the Amazon AWS portal.
Web Portal	A prototypal portal is available
Web Application	The current prototype supports proof-of-concept applications for running workflows
Browser	Most modern browsers supported

Component	Status
Authentication Service	Authentication service not yet available
Authorizer	Authorization not yet available
Logger	Some actions are logged using Amazon and Tomcat logging functionalities

6.4 First Prototype: the Metadata Platform

According to the scheduling shown in Figure 15, a first version of the VLab Platform was released on project month 12 and presented at the ECOPOTENTIAL General Meeting in Texel (NL), on 27-29 July 2016. The prototype is composed of the Metadata Platform, providing data discovery and access capabilities, and proofs-of-concepts for handling workflows related to storylines on different protected areas.

An user can access the Metadata Platform (see Figure 19) through a data portal providing discovery capabilities based on geographical coverage, temporal extent, keywords, and data source. The results are listed with basic information. The user can evaluate the result dataset or collection inspecting the full abstract ("More info" button). Finally, the user can access the result dataset or collection.

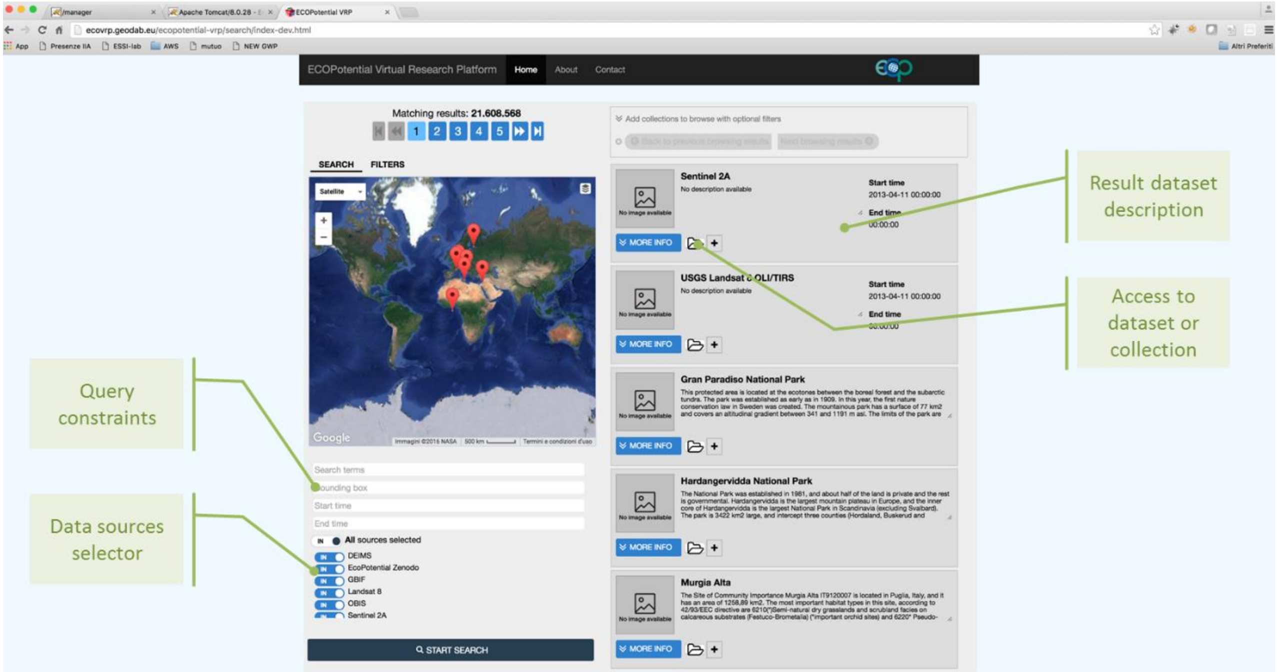

Figure 19 ECOPOTENTIAL VLab prototype: the Metadata Platform functionalities

The proofs-of-concept aim to present a potential access to workflows. The user can access workflows starting from the protected areas (see Figure 20). A simple search field is available to filter potentially interesting protected areas. For each protected area, the portal provides an abstract and links to the relevant ecosystems and storylines.

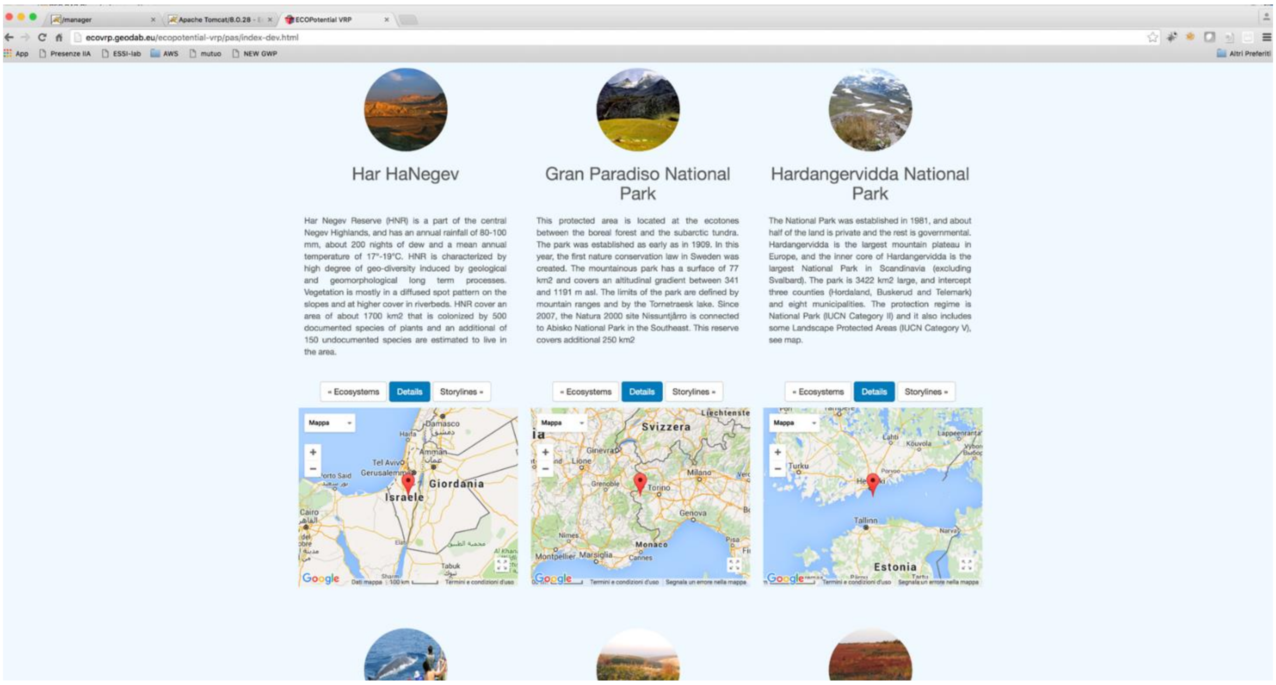

Figure 20 ECOPOTENTIAL VLab prototype: access to protected areas information

Selecting storylines for a protected area, the user can discover the existing workflows for that storyline and protected area (see Figure 21). The workflow is presented through a graphical representation in Business Process Modelling Notation (BPMN). The user can select the desired input data and run the workflow. Currently only a small set of predefined workflows is available as proof-of-concept. However, they are effectively run accessing remote services, showing the feasibility of the proposed approach.

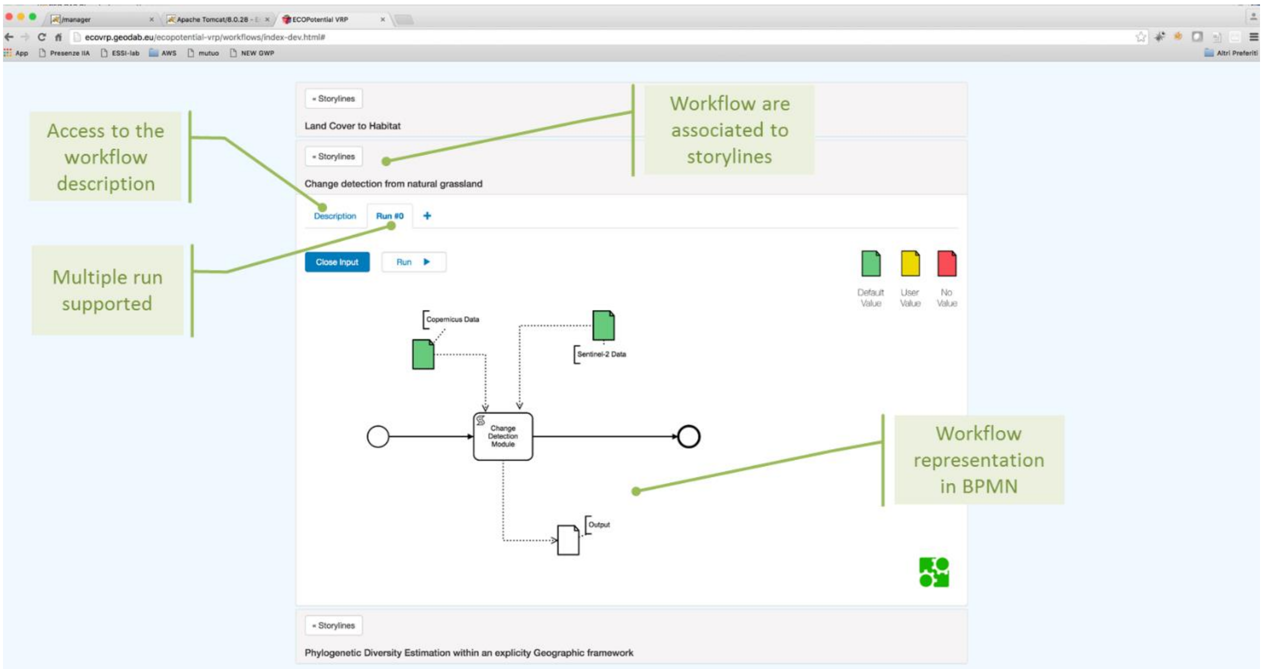

Figure 21 ECOPOTENTIAL VLab prototype: running a workflow

7 DEPLOYMENT

7.1 Deployment plan

As stated in the ECOPOTENTIAL DoW, the ECOPOTENTIAL Virtual Laboratory Platform will be deployed and operated in a cloud-based infrastructure (IaaS). The objective is to support multiple deployment environments.

8 REFERENCES

[1] IEEE, SWEBOK v3 - Guide to the Software Engineering Body of Knowledge, P. Bourque and R. E. Fairley, Eds., IEEE, 2014.

[2] European Commission, "TOPIC : Making Earth Observation and Monitoring Data usable for ecosystem modelling and services," [Online]. Available: https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/378-sc5-16-2014.html. [Accessed 23 May 2016].

[3] European Commission, *Horizon 2020 Work Programme 2014-2015 12. Climate action, environment, resource efficiency and raw materials (revised),* 2015.

[4] GEO, "The GEOSS Data Sharing Principles," 01 October 2015. [Online]. Available: https://www.earthobservations.org/geoss_dsp.shtml.

[5] European Commission, *Grant Agreement number - 641762 - ECOPOTENTIAL,* 2015.

[6] European Commission, "Guidelines on Data Management in Horizon 2020," 15 February 2016. [Online]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf. [Accessed 17 June 2016].

[7] ISO, "Geographic information — Reference model — Part 1: Fundamentals," 2014.

[8] European Parliament; European Council, *Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE),* 2007.

[9] C. M. Bonina, "New business model and the value of open data: definition, challenges and opportunities," 2013.

[10] "Open Definition," [Online]. Available: http://opendefinition.org/od/.

[11] C. Moedas, "Open Innovation, Open Science, Open to the World," 22 June 2015. [Online]. Available: http://europa.eu/rapid/press-release_SPEECH-15-5243_en.htm. [Accessed 23 May 2016].

[12] EGI, "Open Science Commons," EGI, 2015.

[13] ICSU-IAP-ISSC-TWAS, "Open data in a Big Data World," 2015.

[14] Open Knowledge International, "Vision and Values," [Online]. Available: https://okfn.org/about/vision-and-values/. [Accessed 23 May 2016].

[15] A. Carusi and T. Reimer, "Virtual Research Environment Collaborative Landscape Study," 2010. [Online]. Available: https://spiral.imperial.ac.uk/bitstream/10044/1/18568/2/vrelandscapereport.pdf. [Accessed 23 May 2016].

[16] L. Candela, D. Castelli and P. Pagano, "Virtual Research Environments: an Overview and a Research Agenda," *Data Science Journal,* vol. 12, 2013.

[17] A. Cox, "What are communities of practice? A comparative review of four seminal works," *Journal of Information Science,* vol. 31, no. 6, pp. 527-540, 2005.

[18] E. C. Wenger, R. McDermott and W. C. Snyder, Cultivating Communities of Practice: A Guide to Managing Knowledge, Cambridge, MA: Harvard Business School Press, 2002.

[19] P. Oreizy, *Open Architecture Software: A Flexible Approach to Decentralized Software Evolution,* 2000.

[20] M. Jamshidi, System of Systems Engineering - Innovations for the 21st Century, John Wiley & Sons, 2009.

[21] Department of Homeland Security, "The System of Systems Approach for Interoperable Communications".

[22] S. Nativi, M. Craglia and J. Pearlman, "Earth Science Infrastructures Interoperability: The Brokering Approach," *IEEE JSTARS,* vol. 6, no. 3, pp. 1118-1129, 2013.

[23] S. Nativi, D. B. Khalsa Siri Jodha, M. Craglia, J. Pearlman, P. Mazzetti and R. Rew, "The Brokering Approach for Earth Science Cyberinfrastructure," 2011.

[24] P. Zhao, Geospatial Web Services: Advances in Information Interoperability: Advances in Information Interoperability, IGI Global, 2010.

[25] P. Mell and T. Grance, "The NIST Definition of Cloud Computing," 2011.

[26] ISO/IEEE, "Systems and software engineering - Architecture description," 2011.

[27] N. Rozanski and E. Woods, "Software Systems Architecture: Working with Stakeholders Using Viewpoints and Perspectives," Addison-Wesley, 2011.

[28] ISO/IEC, "Information technology — Open Distributed Processing - Reference model: Overview," ISO, 1998.

[29] ECOPOTENTIAL Consortium, "Description of Work," 2015.

[30] ECOPOTENTIAL Consortium, "Data Management Plan," 2016.

[31] M. Santoro, P. Mazzetti, S. Nativi, C. Fugazza, C. Granell and L. Diaz, "Methodologies for Augmented Discovery of Geospatial Resources," in *Discovery of Geospatial Resources: Methodologies, Technologies, and Emergent Applications*, IGI Global, 2012.

[32] M. Holcombe, Running an Agile Software Development Project, John Wiley & Sons, 2008.

[33] K. Beck, M. Beedle, A. van Bennekum, A. Cockburn, W. Cunningham, M. Fowler, J. Grenning, J. Highsmith, A. Hunt, R. Jeffries, J. Kern, B. Marick, M. C. Martin, S. Mellor, K. Schwaber, J. Sutherland and D. Thomas, "Manifesto for Agile Software Development," [Online]. Available: http://agilemanifesto.org/.

[34] ConnectinGEO Consortium, "User Feedback system," [Online]. Available: http://connectingeo.net/UserFeedbackSystem.htm. [Accessed 11 July 2016].

[35] GEO, "GEOSS Data Management Principles," 01 October 2015. [Online]. Available: http://www.earthobservations.org/documents/dswg/201504_data_management_principles_long_final.pdf.

[36] GEOSS-IIB, "GEOSS requirements, functions, architecture," [Online]. Available: ftp://ftp.earthobservations.org/201505_Work_Plan_Symposium/presentations/.

9 ANNEX A: GEOSS DATA SHARING PRINCIPLES

The following data sharing principles were adopted by GEOSS.

DSP1. There will be **full and open exchange** of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;

DSP2. All shared data, metadata and products will be made available with **minimum time delay** and at **minimum cost**;

DSP3. All shared data, metadata and products being **free of charge or no more than cost of reproduction** will be encouraged for research and education.

10 ANNEX B: GEOSS DATA MANAGEMENT PRINCIPLES

The following management principles were introduced by the GEOSS IIB (Infrastructure Interoperability Board) and will be adopted at the next GEO Plenary in Mexico City (November 2015).

Discoverability

DMP-1. Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions, including licenses, will be clearly indicated.

Accessibility

DMP-2. Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.

Usability

DMP-3. Data will be structured using encodings that are widely accepted in the target user community and aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.

DMP-4. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.

DMP-5. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain.

DMP-6. Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

Preservation

DMP-7. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.

DMP-8. Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.

Curation

DMP-9. Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate; where applicable this shall follow established and agreed procedures.

DMP-10. Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

11 ANNEX C: GEOSS ARCHITECTURE PRINCIPLES

The following architectural principles were introduced by the GEOSS IIB (Infrastructure Interoperability Board) and are under discussion.

- Given the nature of a "system of Systems" it was recognized that the **success would depend on building interoperability among the different and autonomous systems.**
- As the basis for evolution and ensure interoperability with relevant research and policy-driven data infrastructures
    - **Openness**
    - **Effectiveness**
    - **Flexibility**
    - **Sustainability**
    - **Reliability**
- Support the implementation of the **Data Sharing and Management principles.**