Report on "5.2.2: iCUPE infrastructure (ver 2)"

Paolo Mazzetti (Consiglio Nazionale delle Ricerche); Steffen M. Noe (Estonian University of Life Sciences)

Florence, Italy, 11 September 2020

WP5: Data provision, interoperability and facilitation of data and services

Task 5.2: Appliance of ERA-PLANET principles and key enabling technologies for interoperability: infrastructure design and development (horizontal) / Deliverable 5.2.2: iCUPE infrastructure (ver 2)

Version 1

Executive Summary

The iCUPE project is an ERA-PLANET Transnational Project aiming at an integrative and comprehensive understanding on polar environments. To achieve such objective, the open and interoperable access to data and generation of knowledge is assured by an ERA-PLANET Knowledge Platform, fully integrated with GEOSS, with functionalities specifically tailored to the iCUPE requirements. The concept of the ERA-PLANET Knowledge Platform stems from the need of lowering barriers for both end-users – to easily access the outcomes of models for knowledge generation – and model developers – to easily publish and share their models. Therefore, it aims not only to data sharing but more generally to provide support to multidisciplinary communities-of-practice for providing knowledge for evidence-based policy and informed decision-making.

The architecture of the ERA-PLANET Knowledge Platform is based on a set of principles currently shared in the scientific research communities, with reference to the GEO intergovernmental initiative, including GEOSS Data Sharing Principles, GEOSS Data Management Principles and GEOSS Architecture Principles. Moreover, since iCUPE, as part of the ERA-PLANET network, participates in the Horizon 2020 pilot action on open access to research data, the activities of the iCUPE Consortium for the definition of the iCUPE Data Management Plan are a fundamental input for the architecture of the ERA-PLANET Knowledge Platform.

The design of the ERA-PLANET Knowledge Platform puts its basis on past experiences in building System of Systems through a brokering approach, and in the development of platforms for knowledge generation, with a specific reference to the Virtual Laboratory developed in the H2020 ECOPOTENTIAL project.

The development adopts a loosely-coupled integration of mature technologies and tools, including Cloud technologies for Infrastructure-as-a-Sevice and Platform-as-a-Service functionalities, and virtualization for building technology-neutral containers.

The integration of new tools in the Knowledge Platform 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, iCUPE adopts an agile methodology allowing rapid development in response to new requirements. It will have yearly iterations with fixed objectives for demonstration in reviews and events.

Table of content

List of Figures

1 Introduction

This document describes the system architecture of the iCUPE instance of the ECOPOTENTIAL Knowledge Platform: a service-based platform for a virtual (i.e. online distributed) and open (i.e. accessible) management and execution of models generating Essential Variables and Indicators.

The present version bases on: (a) the general context of cyberinfrastructures and Virtual Research Environments (VRE) in multidisciplinary science, with particular reference to the ECOPOTENTIAL Virtual Laboratory, (b) the requirements from the *H2020-SC5-2015-one-stage* call on the *SC5-15-2015 - Strengthening the European Research Area in the domain of Earth Observation* topic of the Horizon 2020 Societal Challenge 5, and (c) specific requirements and constraints collected during the preparation of the iCUPE Data Management Plan (DMP) d) the changes in GEO and EuroGEO occurred in 2018 and 2019.

The iCUPE Knowledge Platform is an instance of a more general ERA-PLANET Knowledge Platform (KP), conceived as an enhancement of the ECOPOTENTIAL Virtual Laboratory (VLab), with a specific focus on knowledge management and generation, and a broader scope to multidisciplinarity to address policy goals from International treaties and initiatives. Thus, some architectural principles and solutions comes from the design and development of the ECOPOTENTIAL VLab. For readers' convenience, to avoid continuous reference to external documents, the relevant information on the ECOPOTENTIAL VLab system architecture, documented in the ECOPOTENTIAL deliverable D10.1 [1] is reported in the present document with the necessary changes for ERA-PLANET and iCUPE.

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*" [2]. 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 the present Introduction, a second section focuses on the objectives and rationale behind the project, clarifying the main relevant concepts for iCUPE, such as Open Data, Knowledge and Science, and providing an operational definition of the ERA-PLANET Knowledge Platform.

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

The fourth section describes the iCUPE architectural principles, focusing specifically on the need of loosely coupled applications, and on the brokering approach which is at the core of the ERA-PLANET Knowledge Platform concept.

The fifth section describes the KP 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 iCUPE project, and the sixth and final section reports the deployment plan and achievements at project-month 12.

The present release of the deliverable is based on the first release including a dedicated section (§2.2.1) describing the major changes in the GEO context affecting topics relevant to the design and development of the iCUPE Knowledge Platform, updates based on such new context, and a final section reporting on the main implementation. It also provide updates about the integration with GEOSS.

2 Rationale and main concepts

2.1 The iCUPE main objective

The iCUPE project was proposed as a response to the call for Transnational Projects on the "Polar areas and natural resources" strand, launched by the ERA-PLANET Consortium on September 2016. ERA-PLANET is the project co-funded under the ERA-NET-Cofund scheme of the Horizon 2020 EU Framework Programme for Research and Innovation (H2020). It was proposed in response to the *H2020-SC5-2015-one-stage* call on the *SC5-15-2015 - Strengthening the European Research Area in the domain of Earth Observation topic* aiming at facilitating the generation of evidence-based policies and informed decision-making. The key statement of the call reports the main requirement [3]:

> *Decision makers require access to the information they need, when they need it, and in a format they can use.*

More specifically, the "Polar areas and natural resources" strand focuses on achieving one of the most important ERA-PLANET objectives to address the main requirement [4]:

> *[ERA-PLANET] will provide more accurate, comprehensive and authoritative information to policy and decision-makers in key societal benefit areas, such as […] Polar areas and Natural resources. […]*

As part of the activities to achieve this goal, iCUPE proposed an operational objective [5]:

> *to optimize data streams and foster interoperability and data sharing for polar observations, in a GEOSS and Copernicus perspectives.*

Aiming to

> *deliver novel data products, metrics and indicators to the stakeholders concerning the environmental status, availability and extraction of natural resources in the polar areas. These data, metrics and indicators will be targeted to identified stakeholders. They will be useful for policy development and for improving and clearly communicating our multidisciplinary understanding of status of the polar environment and pollution dynamics in the future. The knowledge generated is relevant to the general population, policy makers and scientists.*

The present document describes the system architecture for reaching this operational objective.

2.2 The ERA-PLANET context and conditions

The ERA-PLANET context poses some significant conditions on how the operational objective must be fulfilled:

C1. Integration with GEO, Copernicus and ESA activities concerning data sharing and knowledge generation. (The *H2020-SC5-2015-one-stage* call specifically required to address the coherence of European participation within GEO and provide a research and innovation component to the Copernicus programme [4].)

C2. Alignment with major open data initiatives. (The ERA-PLANET deliverable D4.5 "Data Management Plan" provides the Data Management Principles for ERA-PLANET presenting the general context that ERA-PLANET Transnational Projects live in [6]. Specific iCUPE Data Management Principles and iCUPE Data Management Plan will be periodically updated as iCUPE deliverables D5.1.x and D5.3.x [5].)

C3. Adoption of an ERA-PLANET common set of Key Enabling Technologies (KET). (The call for Transnational Projects explicitly asked for a plan of adoption of KETs including the GEOSS Platform components.)

2.2.1 Changes in the GEO context

In the period since the start of the iCUPE project and the release of the present deliverable (June 2020) some important changes have occurred in the GEO context:

G1. The new GEO Secretariat Director Gilberto Camara, in charge since July 2018, proposed his vision of a Results-Oriented GEOSS with a special focus on knowledge management and sharing for science reproducibility and replicability through the building of a GEO Knowledge Hub.

G2. In support of the Director's vision, the GEO Secretariat established an Expert Advisory Group (EAG)[1] to advice about the details and operational objective for the new GEOSS and the GEO Knowledge Hub. The EAG finished its activity on April 2019 releasing a strategic document entitled "Results-Oriented GEOSS: A framework for transforming Earth observation data to knowledge for decision making" for discussion to the GEO Executive Committee of July 2019. In this document, the EAG proposes "*an instrumental framework to advance a Results-Oriented GEOSS that includes foundational pillars, goals, objectives, and actions to transform the current data focused GEOSS to a knowledge-based GEOSS delivering decision-ready products and services*."

G3. In 2018-2019, the Work Programme 2020-2022 was prepared and finally approved at the GEO XVI Plenary on November 2019. Following the Director's vision, the EAG suggestions and Programme Board indications, the Work Programme 2020-2022 includes major changes in the structure and objectives of the Foundational Tasks, specifically reflecting the shift towards knowledge generation and sharing.

G4. The EuroGEO (formerly EuroGEOSS) initiative for the European contribution to GEOSS, launched by the European GEO Caucus at the GEO Week 2017 (October 2017), underwent significant changes and launched specific activities and projects. In particular:

a. An EuroGEO Coordination group with representatives from the member states was established;
b. EuroGEO organized its activities in a set of thematic EuroGEO Action Groups on: Applications for Agriculture/Food, Applications for general Land use/land coverage, Urban applications (including urban air quality and urban health), Applications for Disaster Resilience, Applications for Biodiversity & Ecosystems, Marine applications, Applications for Climate, Applications for Atmosphere, Applications for Energy;
c. A dedicated call was launched by the European Union under the Horizon 2020 Framework Programme to support the implementation of EuroGEO showcases. The winning project called E-SHAPE started on May 2019;
d. The former GEO European Project Workshop series was re-nominated as EuroGEO Workshop since the 2018 edition to reflect the focus on the EuroGEO activities;
e. On February 2019, the European Commission Directorates DG-GROW, DG-RTD, and DG-JRC jointly launched the EuroGEO Sprint-to-Ministerial initiative with a call for proposals for EuroGEO showcases to be presented at the GEO Week 2019;
f. On July 2019, following the indication of GEO XV Plenary and in line with the decision of analogous regional initiatives, the original EuroGEOSS name was changed to EuroGEO to identify the European regional GEO, assuming that a single GEOSS will exist;

It worths noting that the iCUPE instance of the ERA-PLANET Knowledge Platform is kept aligned with the previous changes. In particular, many iCUPE Consortium key persons and more generally ERA-PLANET Consortium key persons play central role in GEO and EuroGEO initiatives, tasks and boards. The

[1] One of the authors (P. Mazzetti) of the present deliverable was a member of the EAG.

ERA-PLANET Knowledge Platform, as its ancestor the ECOPOTENTIAL VLab, have been designed and developed taking into account the GEO architectural framework from the GEOSS EVOLVE Initiative. Clearly, the ERA-PLANET Knowledge Platform perfectly fits as a contribution to the new focus of GEOSS in knowledge generation and sharing. More specifically, the VLab/KP is a central component of a showcase for the EuroGEOSS Sprint-to-Ministerial on Land Cover Change and Land Degradation and it is proposed for the implementation of some E-SHAPE showcases. It worths noting that the Land Cover Change and Land Degradation, based on the ERA-PLANET Knowledge Base, was selected by the European Commission to demonstrate the EuroGEO contribution to GEO duirng the GEO XVI Plenary.

2.3 Geospatial data, information and knowledge in ERA-PLANET: roles and issues

The generation of knowledge for decision-making about resource efficiency and environmental management 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 producers* who acquire observations (e.g. through sensors);
- *Knowledge providers* who 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 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;

- *Knowledge providers* are mostly focused on models and algorithms that implement them. Several frameworks exist which address different needs and knowledge generation approaches. They range from general-purpose programming frameworks (Java, Python, R, etc.) to full platforms dedicated to specific modelling approaches (e.g. for Bayesian Belief Networks, Neural Networks, etc.)
- *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 SC5-15-2015* topic explicitly mentions this issue saying that an expected impact is a "*significant improvement of shared Earth Observation architectural components and related information infrastructure, improved, open and unrestricted data sharing across borders and disciplines, and interoperability amongst observational, modelling, data assimilation and prediction systems to maximise value and benefits of Earth observation investments*" [3].

2.4 Open Data in ERA-PLANET

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 iCUPE, the call provides few hints limiting to state that an expected impact is "*improved, open and unrestricted data sharing across borders and disciplines and interoperability amongst observational, modelling, data assimilation and prediction systems*" [3]. This definition clarifies that iCUPE will handle multiple resources typologies.

The main source of information about resource management in iCUPE - and more generally in all the ERA-PLANET Transnational Projects - is the ERA-PLANET deliverable D4.5 summarizing the requirements, mainly from H2020 Open Research Data Pilot, GEOSS and Copernicus in a set of Data Management Principles [6]:

EDMP-1 All data generated in the action must be deposited in a research data repository and made accessible free of charge and at the FAIR conditions described in the DMP;

EDMP-2 All the scientific results generated in the action (e.g. presented in a publication) must be reproducible providing the required data and information about tools and instruments necessary for validation;

EDMP-3 All data generated in the action, which are relevant, directly or indirectly, for information to policy and decision-makers in key societal benefit areas must be accessible through GEOSS and Copernicus at the conditions described in the DMP and in compliance with GEOSS-DSP and GEOSS-DMP;

iCUPE will document its Data Management Plan (DMP) following the ERA-PLANET Data Management Principles in multiple release of the deliverable D0.2.

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 the ERA-PLANET Knowledge Platform

The ERA-PLANET Consortium proposed the realization of an ERA-PLANET Knowledge Platform as the answer to the call request for providing "more accurate, comprehensive and authoritative information to policy and decision-makers in key societal benefit areas". Since many requirements are in common with the other ERA-PLANET Transnational Project, the iCUPE Consortia recognized that the general ERA-PLANET objective of delivering to decision makers "the information they need, when they need it, and in a format they can use" can be achieved through the implementation of a ERA-PLANET Knowledge Platform tailored to knowledge generation and management, and customized for each Transnational Project. To this aim we provide the following definition:

> The ERA-PLANET Knowledge Platform is a virtual environment enabling knowledge generation from heterogeneous data sources.
>
> The iCUPE Knowledge Platform is an instance of the ERA-PLANET Knowledge Platform customized for informed decision-making in support of decision making for the management Polar areas and Natural resources.

In the definition above:

- *Virtual* means that there is not any necessity to physically centralize resources. For example, the KP provides access to heterogeneous data, but data do not need to be moved from their original site; the KP stores the representations of models, but the algorithm code can be stored in a remote repository and it can be executed by a remote processing service on a cloud. This is specifically important in the

EuroGEO context where the use of European resources for data (e.g. Sentinel data series) and processing services (e.g. Copernicus DIAS) is encouraged.

- *Environment* means that the user experience is that of a controlled space where he/she can operate through a dedicated user interface.
- *Knowledge generation* refers to the full cycle of transition from data to knowledge, including the generation of Essential Variables and the proper indicators for assessing policy targets.

3 ERA-PLANET architecture principles

To comply with ERA-PLANET requirements, ERA-PLANET introduced a set of architecture and interoperability principles to facilitate discovery, access, (re-)use, and preservation of data and algorithms implementing models for knowledge generation:

AP1. To build the ERA-PLANET data and services infrastructure on the existing and under development digital systems –noticeably, the digital systems identified in the thematic workpackages (WP4, WP5, WP6).

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 to the ERA-PLANET platform for discovering, accessing, and using ERA-PLANET services –for interoperability to GEOSS, Copernicus, and other EC-funded programmes.

AP4. To comply with the GEOSS Architecture Principles, (see Annex C). Also taking into account the recent evolution towards a Results – Oriented GEOSS (see §2.2.1).

AP5. To comply with the ERA-PLANET Data Management Principles.

AP6. To comply with the ERA-PLANET Key Enabling Technologies (KET) guidelines

4 ERA-PLANET design principles

The design principles translate the architecture principles in guidelines for the design of the ERA-PLANET Knowledge Platform.

4.1 Open Software Architectures

The design of the ERA-PLANET Knowledge Platform 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, ERA-PLANET conceives a Knowledge Platform 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 KP 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 KP must be also able to evolve in response to those changes. Indeed, although the KP requirements can become clear during the course of the ERA-PLANET and iCUPE project, in order to support the sustainability of outcomes, it is necessary to assure that the KP architecture and implementation can (easily) evolve.

Software evolution has been the subject of several research works in the past (Table 1). A first classification [16] 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 ERA-PLANET KP for several reasons: a) a KP is not fully based on software which is under control of a single organization (e.g. models may be developed by external organizations); b) even the ERA-PLANET 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 ERA-PLANET Consortium could achieve the role of central authority, it exists only until the end of the project, whereas the sustainability of KP must be considered also beyond the ERA-PLANET and iCUPE projects 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 ERA-PLANET 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 ERA-PLANET 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 **ERA-PLANET Knowledge Platform adopts 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 ERA-PLANET and iCUPE. To address interoperability the ERA-PLANET 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, 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* [17]. It is evident that this definition fits well in the iCUPE context where sub-systems like the INSPIRE infrastructure, Copernicus core and downstream services are clearly out of control of the ERA-PLANET Consortium, and even from possible future exploitation scenarios.

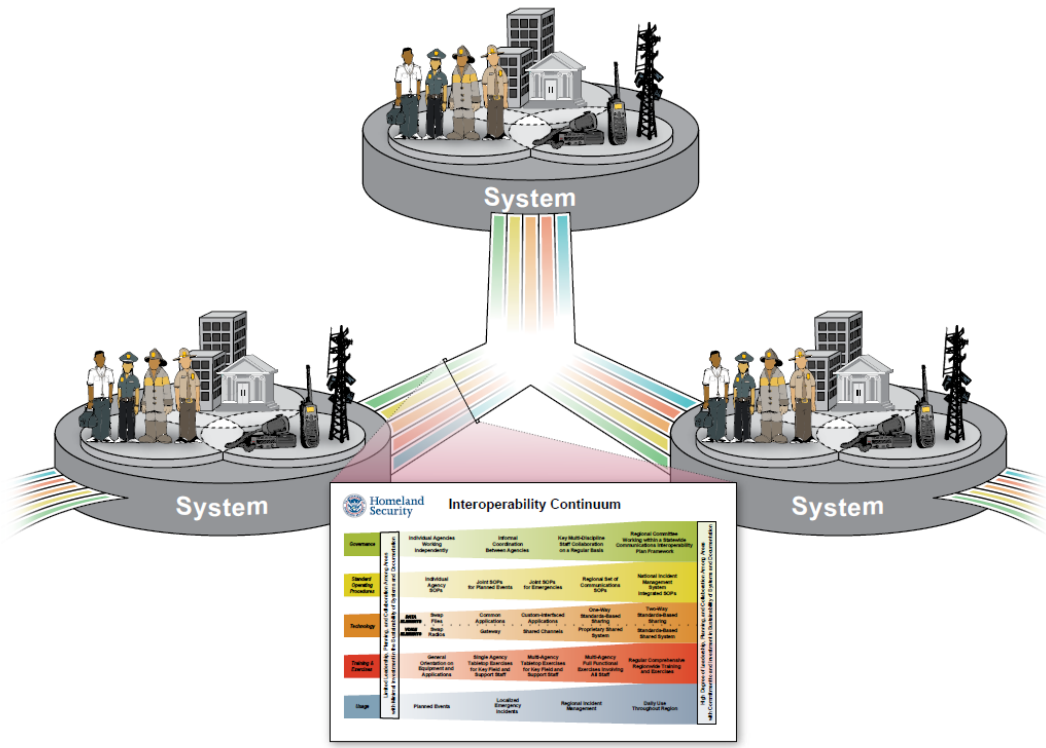

Figure 2 System of Systems in Practice – from [18]

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* [19] [20], 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" [21]. 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" [21]. 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[2] 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.

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 ERA-PLANET, the choice of brokered architectures is fully justified by two main reasons:

a) There are several data sources of interest for ERA-PLANET 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)[3]
    b. In the meteo-ocean community, the UNIDATA THREDDS Data Server (TDS)[4] is widely adopted

[2] https://rd-alliance.org/groups/brokering-governance.html
[3] http://www.tdwg.org
[4] www.unidata.ucar.edu/software/thredds/current/tds/

   c. Many Open Data communities share the CKAN[5] technology for implementing data portals.

b) ERA-PLANET 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. ERA-PLANET 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.

[5] http://www.ckan.org

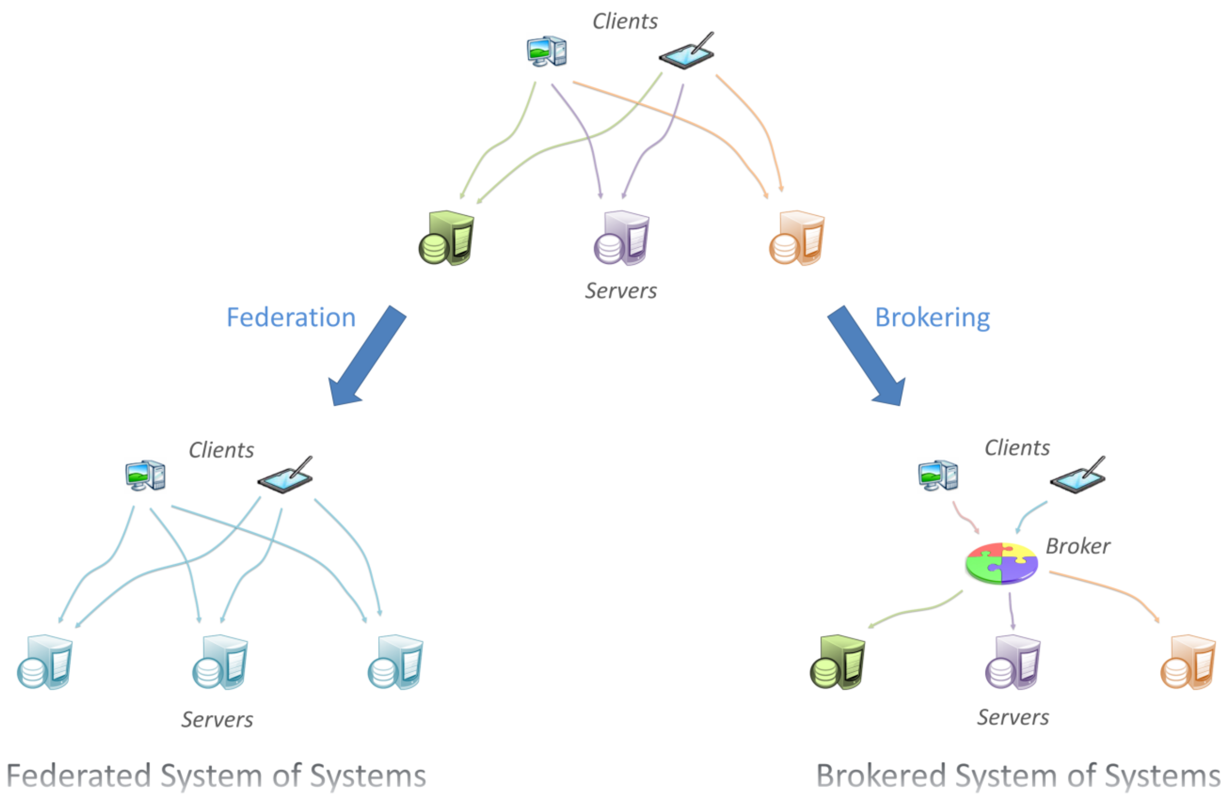

Figure 3 Federated vs. Brokered Architectures for Systems of Systems

4.3 ERA-PLANET 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*" [22].

The cloud model includes three different kinds of services [22]:

- *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 can 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 widens 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 ERA-PLANET, where there is no 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 Knowledge Platform with their own browser.

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

The KP 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

ERA-PLANET 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 ERA-PLANET 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 ERA-PLANET 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 ERA-PLANET design principles

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

DP1. ERA-PLANET Knowledge Platform adopts an Open Software Architecture

DP2. ERA-PLANET Knowledge Platform is developed integrating and adapting existing software solutions

DP3. ERA-PLANET Knowledge Platform adopts a Decentralized Software Evolution

DP4. ERA-PLANET Knowledge Platform is made of software components interacting through (low-level) APIs

DP5. ERA-PLANET Knowledge Platform relies on a common infrastructure of a brokered System of Systems for data interoperability

DP6. ERA-PLANET Knowledge Platform exposes a set of (high-level) APIs for interaction with the external environment

DP7. ERA-PLANET Knowledge Platform is accessible according to the Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) models, for end-users and developers respectively

DP8. ERA-PLANET Knowledge Platform can be deployed either on private infrastructures or commercial or public clouds providing Infrastructure-as-a-Service (IaaS) capabilities.

DP9. ERA-PLANET Knowledge Platform security architecture is orthogonal to the ERA-PLANET Knowledge Platform resource-sharing architecture.

5 ERA-PLANET Knowledge Platform 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" [23]. 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" [24].

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*" [24].

The following paragraphs provide the ERA-PLANET Knowledge Platform description according to the following main views adopted in the ISO Reference Model for Open Distributed Processing (RM-ODP) [25]:

- 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;
[25]

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

ERA-PLANET identifies a set of Actors, which is a set of user categories involved in: a) the setup and operation of the Knowledge Platform, b) the use of Knowledge Platform resources, and finally, c) the use of applications based on the Knowledge Platform. They are

Actor	Acronym	Description
Knowledge Platform Provider	KP Provider	The KP Provider is the person/organization that provides the KP capacities.
Knowledge Platform Administrator	KP Admin	The KP Admin is the person who manages a Knowledge Platform configuring it for KP users and providing support.

Knowledge Platform End User	KP End User	The KP End User is the person who gets value from the KP (e.g. policy-maker, decision-maker)
Knowledge Platform App Developer	KP App Developer	The KP App Developer is an intermediate user, a person who develops and manages applications based on the KP APIs.
Knowledge Platform Consumer	KP Consumer	A KP Consumer is a person who makes use of KP capabilities, which is either a KP End User or a KP App Developer.

Table 2 Description of ERA-PLANET Knowledge Platform actors

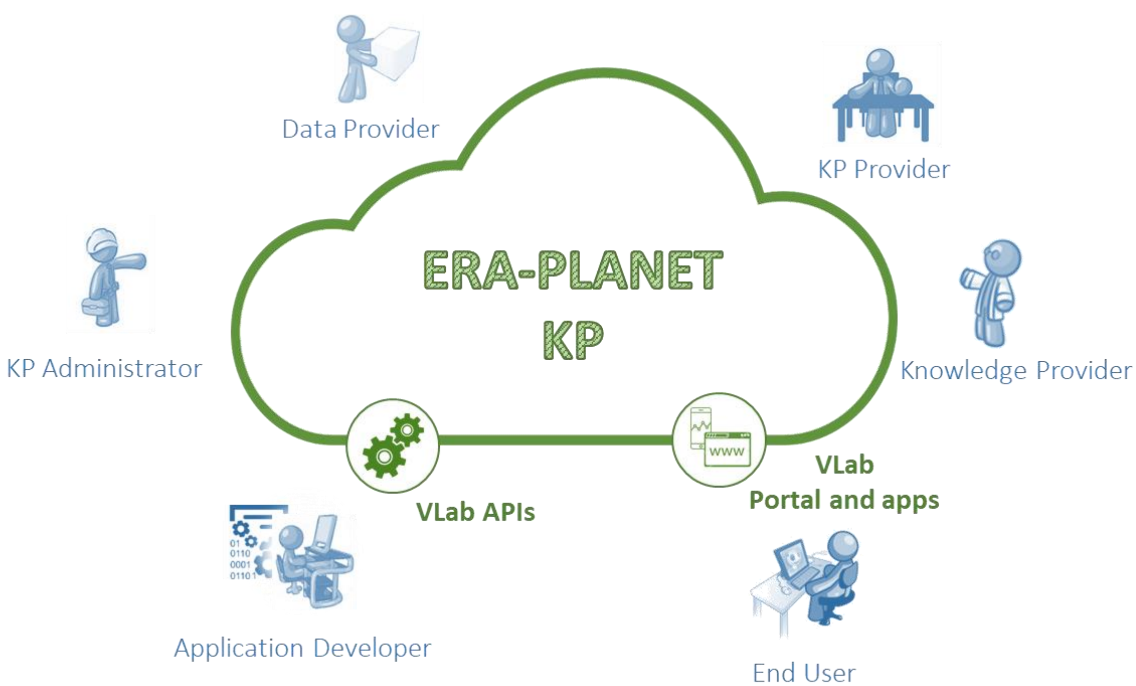

Figure 4 Knowledge Platform actors (in blue)

5.2.2 User scenarios and requirements

The General scenario: from Data to Knowledge

In the recent years, as part of the activities of international organizations like the United Nations (UN), or as defined by international agreements, several policy goals to be achieved in a defined timespan have been set. They include the 17 UN Sustainable Development Goals (SDGs), the objectives of the Sendai Framework on Disasters, the objectives of the Conference of Parties 2015 on Climate (COP21) and so on. The achievement of these policy goals can be measured in respect of specific policy targets. The assessment of targets, and in general the definition of possible actions towards their fulfillment require informed decision-making. Therefore, policy-makers are asking the scientific community to provide the necessary knowledge for evidence-based decision-making. This results in the necessity of extracting knowledge from the big amount of collected data including EO and in-situ observations, and from the available socio-economic information.

This transition from Data to Knowledge require filling the wide gap between acquired data, and policy targets and goals. A procedure proposed and elaborated as a result of recent research and innovation projects consist in adopting a step-by-step approach. Acquired data can be processed to generate information in form of Essential Variables, defined as the physical parameters necessary to describe the status of the Earth system in a domain (therefore we have Essential Variables for each relevant scientific domain). Then, Essential Variables can be processed to summarize the knowledge in one or more Indicators whose value is related to a specific target. Comparing the value of the Indicators from a real or simulated situation against the target it is possible to assess and evaluate the fulfilment of a specific policy goal.

The ERA-PLANET KP aims at supporting the Data to Knowledge transition as the general scenario. In iCUPE this is specifically focused on supporting the assessment of Policy Goals (see Figure 5). This means that iCUPE aims at lowering barriers to the implementation of the proposed transition from Data to Knowledge for achieving policy goals. Specific challenges are:

- Big Data: the ERA-PLANET KP must address Big Data issues specifically handling big volume challenges (big datasets or large number of small datasets) and large variety challenges (different content, communication protocols, formats, Coordinate Reference Systems, etc.).
- Knowledge generation: the ERA-PLANET KP must support different methods for generating knowledge, including integrated modeling, machine learning and deep learning, etc.

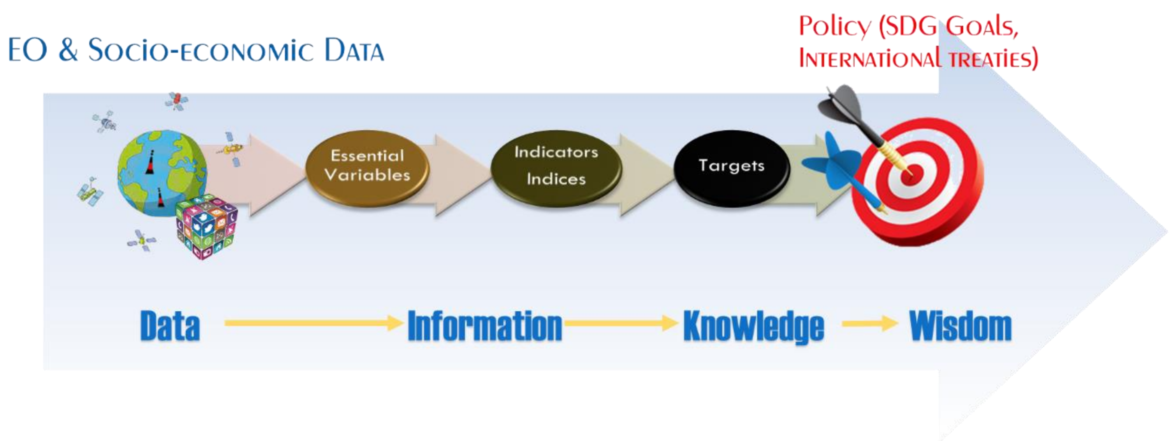

Figure 5 The ERA-PLANET general scenario: from Data to Knowledge

Detailed scenarios and requirements

The ERA-PLANET User Requirements are collected from different sources:

a) ERA-PLANET Call text [3]

b) Previous work in relevant initiatives and programmes at national, regional, European and international level (including Copernicus, INSPIRE, GEOSS)

iCUPE specific requirements are collected from other sources:

c) iCUPE Consortium Agreement and DoW [5]

d) Collection of user requirements from iCUPE Thematic Work Package WP4 "Integrating in-situ, satellite and model components for improved environmental assessment"

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

In terms of user requirements, the ERA-PLANET Knowledge Platform was conceived as a resource sharing system. The high-level use-cases are those needed to support the typical resource sharing scenario shown in Figure 6, 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 6 The typical high-level scenario in the resource sharing

In ERA-PLANET the term *resource* encompasses:

- Data
- Semantic assets
- Scientific workflows
- Analytic services

Table 3 reports the main user scenarios for the ERA-PLANET KP and iCUPE instance.

	User Scenario	Description
S1.	Search for datasets	The KP End User: a) searches for available datasets per geographical coverage, temporal extent, keywords, concepts; b) evaluates available datasets through metadata; c) downloads relevant datasets in the preferred format, resolution, etc.
S2.	Search for Policy Goal	The KP End User: a) browses Policy Goals (e.g. UN SDGs); b) chooses one Policy Goal; c) gets available information about the selected Policy Goal (including available Indicators)
S3.	Search for Indicators	The KP End User: a) browses Indicators; b) chooses one Indicator; c) gets available information about the selected Indicator (including available Workflows to generate it)
S4.	Publish resources	The KP 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 Policy Goal	The KP End User add a Policy Goal with all the required information to the KP
S6.	Add an Indicator	The KP End User add an Indicator with all the required information to the KP including which Policy Goal it refers to
S7.	Add a Workflow	The KP End User adds a Workflow with all the required information to the KP including which Indicator it refers to, optionally uploading the source/executable code and a web service endpoint
S8.	Run a Workflow	The KP End User: a) browses Policy Goals; b) chooses one Policy Goal; c) browses the available Indicators for that Policy Goal; d) chooses an Indicator; e) browses the available Workflows; f) chooses one Workflow; g) select input datasets available for that Workflow; h) runs the Workflow; i) accesses the result
S9.	Run apps	The KP 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

S10.	Develop a new app	The KP 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 iCUPE marketplace

Table 3 User scenarios for the ERA-PLANET Knowledge Platform

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

User Requirement	Description
Data publishing	The KP Consumer is able to publish single datasets or connect data systems with minimal interoperability agreements
Harmonized access to data	The KP Consumer is able to seamlessly discover and download data from heterogeneous sources
Data harmonization	The KP consumer is able to download data harmonized in terms of format, spatial and temporal coverage, coordinate reference system, resolution.
Scientific workflow publishing	The KP Consumer is able to publish a formal representation of a model
Scientific workflow access	The KP Consumer is able to discover, visualize and run models
Analytic service publishing	The KP Consumer is able to publish a Web service implementing a model
Analytic service run	The KP Consumer is able to run a Web service implementing a model
Semantic enrichment	The KP Consumer is able to use semantic assets for suggestions, multilingual discovery, etc.

Table 4 User requirements for the ERA-PLANET Knowledge Platform

5.2.3 Constraints and assumptions

The main constraint for ERA-PLANET Knowledge Platform and its iCUPE instance is that the architecture must support the ERA-PLANET Data Management Principles and the exposed resources must comply with the iCUPE Data Management Plan.

5.2.4 System Requirements

The ERA-PLANET/iCUPE System Requirements are collected from different sources:

a) Call text [26]
b) iCUPE DoW [5]
c) Elicitation from user requirements (see Section §5.2.1)
d) Data Management Plan [27]
e) Specific requirements from iCUPE Thematic Work Packages WP4 "Integrating in-situ, satellite and model components for improved environmental assessment"

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 §0)
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 §0) and dedicated APIs
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 evaluation	The system provides description of algorithms
FR6	Algorithm 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) Policy Goal b) Indicator
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	Data registration in GEOSS	Data available in the KP 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 iCUPE system requirements

It is worth noting that, at this level of details, no iCUPE specific functionality is requested. The personalization of the platform for iCUPE purposes consists in tailoring the content. For example, the general FR7 requirement (scientific workflow discovery) will be specialized supporting iCUPE concepts like Policy Goal and Indicator for polar areas, which, in turn, will be defined in the Knowledge Base and their access assured by compliance with FR2.

Therefore, the system requirement elicitation confirms that the design of a general-purpose ERA-PLANET KP open to domain specialization, is a valid approach.

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.
[25]

Figure 7 shows the high-level architecture of the ERA-PLANET Knowledge Platform. It includes the following layers:

- **Resource** layer: this layer provides functionalities for publishing, discovery and access resources on heterogeneous data systems.
- **Integration** layer: this layer provides functionalities supporting workflows based on the harmonized resources provided by the lower layer.
- **Application** 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.

This three-layer architecture provides three levels of services:

- **Upstream** services: resource provision services exposed by the Resource Layer
- **Midstream** services: interoperability and processing services exposed by the Integration layer for building end-user applications.
- **Downstream** services: end-user services provided by the Application layer.

It also implements the Separation-of-Concern pattern with different specialized stakeholders focusing on the design and development of the functionalities of each layer:

- **Resource Providers** can focus on the Resource layer to assure that resources are provided with the best level of service possible.
- **Interoperability Experts** can focus on the Integration Layer to address heteroegenity of data and process to expose harmonized resources and facilitate application development.
- **Application Developers** can focus on the application business logic building on harmonized resources.

The KP is accessible at different level through:

- Graphical-User-Interface: an user interface designed for the interaction with the proper human users (e.g. end-users at the Application Layer, or Resource Providers at the Resource Layer).

- Services: services exposing the layer functionalities for machine-to-machine interaction.
- APIs: machine-to-machine interfaces exposing resources to build applications on top of them.

It is expected that each layer will have a predominant way of interaction: GUI at the Application Layer, APIs at the Integration Layer, Services at the Resource Layer.

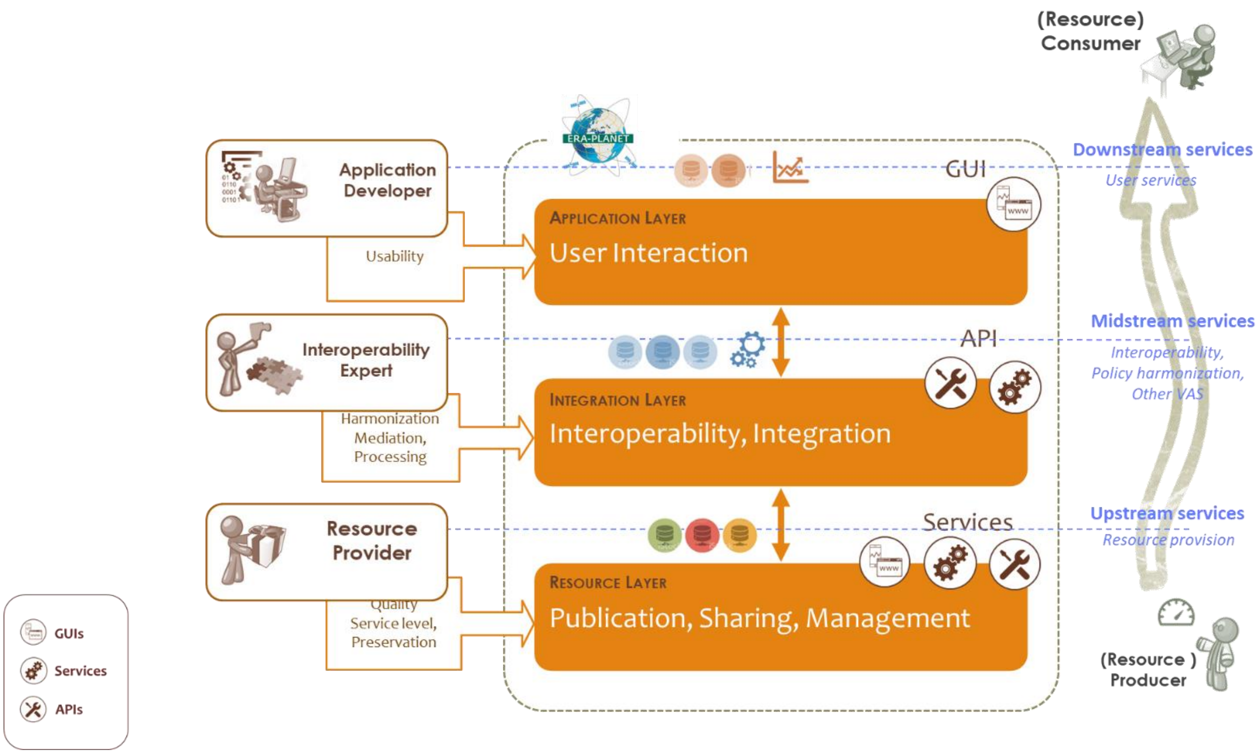

Figure 7 ERA-PLANET KP layered architecture

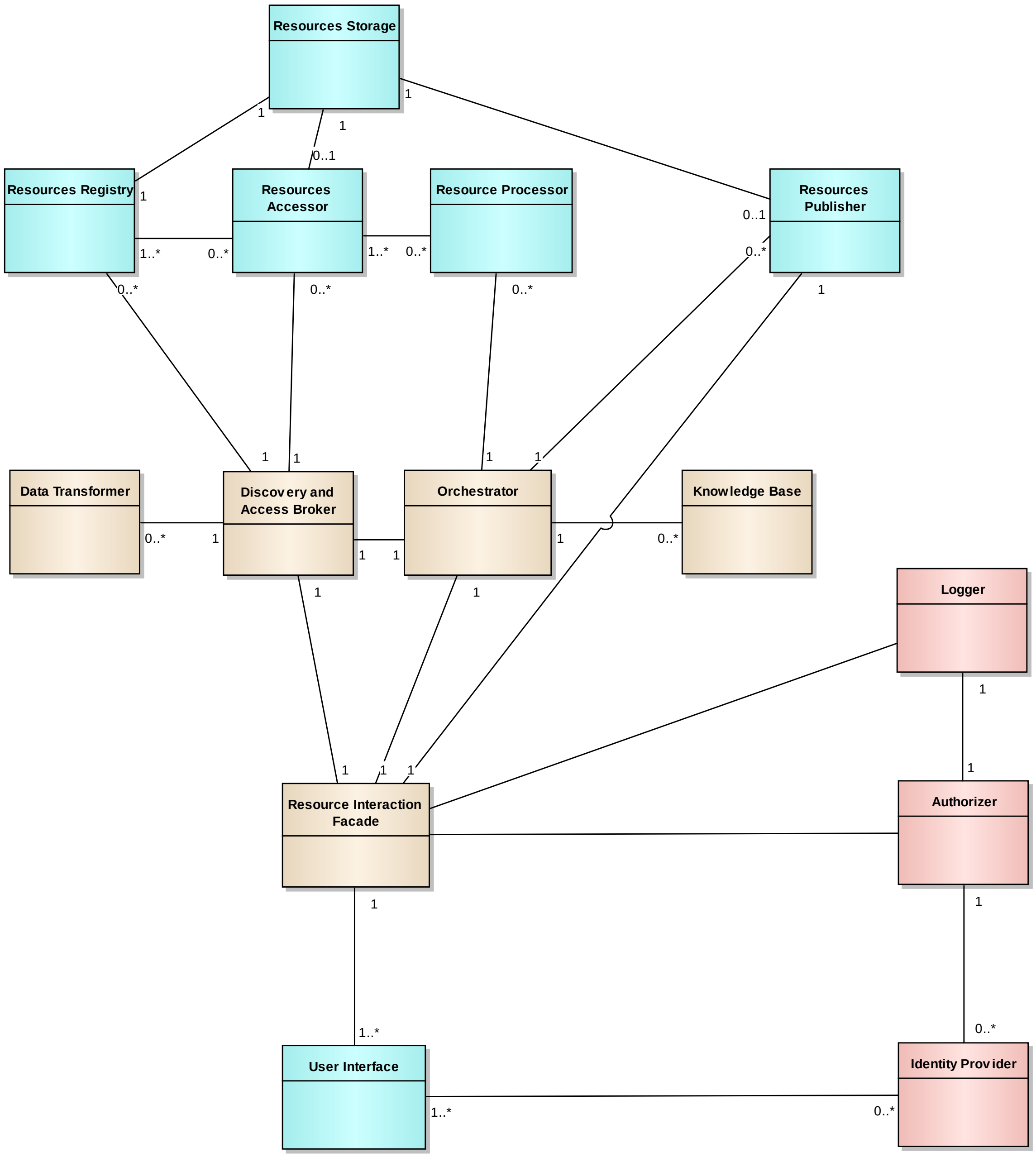

Figure 8 Main components of the ERA-PLANET KP logical architecture (security components in red; external components in blue)

Figure 8 shows the UML class diagram of the main functional components in the ERA-PLANET KP architecture. Components in red are not involved in the resource-sharing functionalities of the iCUPE system; they are either components of the security architecture or ancillary components improving the ERA-PLANET KP overall system capabilities. Components in blue are considered external components and already available as external services, or components (partially) described in other iCUPE documents (e.g. the User Interface is implemented as a Dashboard).

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 Knowledge Platform	FR11, FR12
Resource Processor	It provides processing of resources	FR9
Discovery and Access Broker	It accesses multiple Resources Registries and Resources Accessors providing harmonized discovery and access to heterogeneous sources	FR1, FR2, FR3, FR7 NFR1, NFR8
Orchestrator	Based on user requests, and resource descriptions in the Knowledge bases, it access resources through the Discovery and Access Broker, invokes Resource Processors to generate new resources and publishes them.	FR9, NFR8
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 according to the iCUPE Ontology, to support advanced discovery and processing services	FR2

Resource Interaction Facade	It provides a common and simplified interface to the KP 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 dedicated, apps, development portal and the iCUPE Dashboard	NFR7

Table 6 ERA-PLANET KP main components

From Figure 7 and Table 6 it is evident the core roles of the Orchestrator and the Discovery and Access Broker (DAB). The Orchestrator implements all the requested logic to run workflows on specific location, and data. The DAB harmonizes interaction with multiple resources impacting on many functional and non-functional requirements.

Table 6 shows how the proposed computational architecture addresses all the functional requirements. Of course, it also impacts on some non-functional requirements (mostly related to the adopted layered architectural style). 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 5.6 and 6, 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.
[25]

In ERA-PLANET two aspect of shared information are concerned:

- The *information typology*, which is relevant from an architectural point-of-view to define interoperability challenges
- The *information content*, which is relevant from an architectural point-of-view to refine use-cases

The information typology concerns the nature and characteristics of the resource artifacts that the KP handles. The ERA-PLANET KP manages different kinds of information resources:

- Satellite data
- In-situ data
- Processing Algorithms
- Models/Workflows
- Model results (products)

This classification is important for defining interoperability challenges and solutions.

The information content concerns the semantic level of the resource artifacts that the KP must handles. The ERA-PLANET KP manages information resources at different semantic level. For example:

- Raw data
- Pre-processed data
- Essential Variables
- Knowledge generation models
- Indicators

This classification is important for defining the ERA-PLANET KP support to the Data to Knowledge transition.

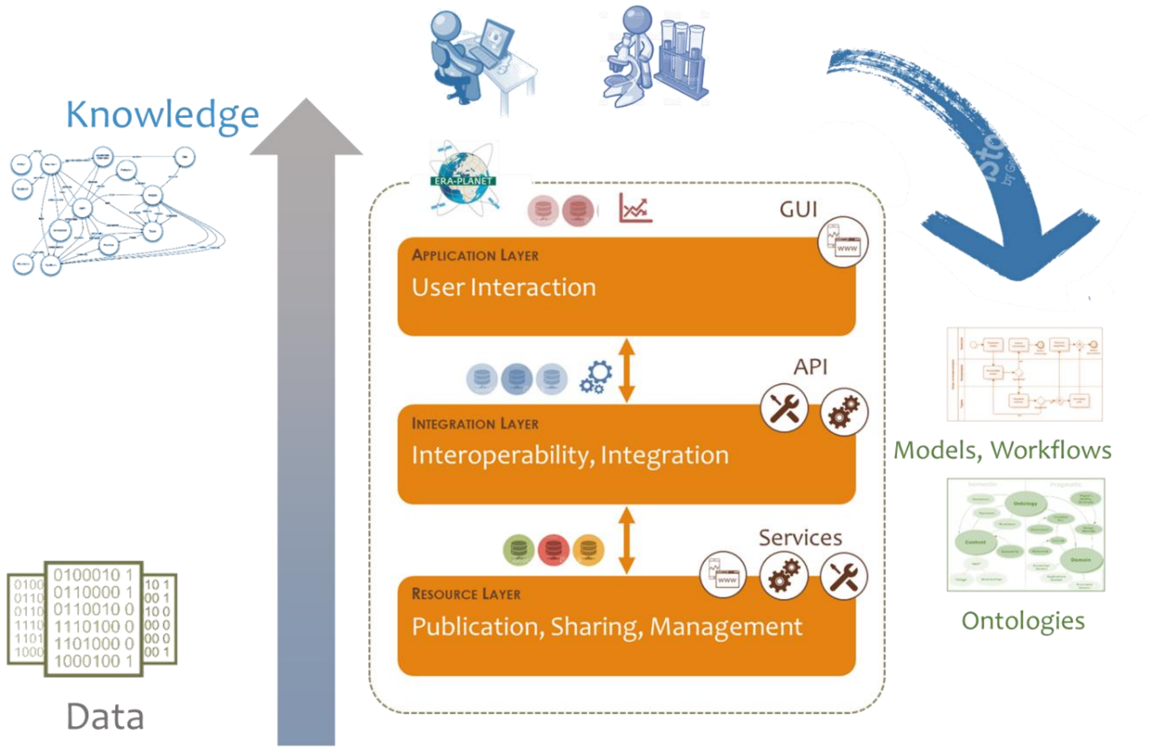

Figure 9 The ERA-PLANET KP architecture and the transition from Data to Knowledge

It is worth noting that the two proposed classifications of resources (typology and content) are not orthogonal. For example, a Model Result could be an Essential Variable, or an Indicator. However, they are both necessary for the definition of the ERA-PLANET KP architecture. Model Result is a concept which is useful for the design of the KP, since it is something that: a) is generated (by a Processing Algorithm), b) should be correctly documented, c) published, d) visualized, etc. The KP architecture must support the Model Result lifecycle. On the other hand, to make the KP useful, and make possible to build applications on top of it, it is necessary to provide a semantic content. Therefore when documenting a Model Result it is necessary to inform that it is either an Essential Variable, or an Indicator, etc.

5.4.1 Resource sharing in iCUPE

As a project finalized to the creation of a Knowledge Platform facilitating the workflows based on heterogeneous resources, the characteristics of information (information typology) handled and shared by the system is a fundamental aspect.

iCUPE addresses two main challenges concerning information handled by the Knowledge Platform:

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

Heterogeneity

As said above, the ERA-PLANET Knowledge Platform aims to facilitate the use of many different kinds of geospatial resources including Satellite data, In-situ data, Processing Algorithms, Models/Workflows, Model results (products).

The heterogeneity challenge is particularly important for data. Indeed, althourgh for some resources like workflows, which are not widely shared yet, it is possible to define and adopt common model to be used in iCUPE, for data this is not possible. Geospatial data comes in many different forms, and iCUPE cannot assume any kind of standardization (see section §2.4). As such, the KP 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 iCUPE, from data acquisition, to processing for knowledge generation and publishing of outcomes.

In general, the KP must be able to handle different service interfaces and metadata/data models for data discovery and access. The final release of the iCUPE DMP will provide the full list of data and data service specifications to be supported. However, already at this stage, it can be foreseen that the ERA-PLANET KP must support all the major de-iure and de-facto standards adopted in the relevant Communities-of-Practice.

The KP must also be able to address interoperability of processing algorithms. Although, at higher level, a model can be described using a common notation (e.g. BPMN), the technologies adopted for algorithm encoding – i.e. the programming framework – and publishing – e.g. as source code on different sharing systems, or as remote service – can vary. The ERA-PLANET KP must support as many systems as possible keeping a high level of technology neutrality in order to lowering entry barriers for developers.

Semantics for improved resource sharing

The KP addresses semantics for improved resource sharing through a query expansion strategy. When a query is submitted to the KP, the KP can ask external semantics services, to resolve keywords, providing "related" terms back. The returned concepts are used as keywords of multiple geospatial queries [28]. 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 8, and Figure 12 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.

5.4.2 Information content: the ERA-PLANET Ontology

The ERA-PLANET KP must support the transition from Data to Knowledge for the assessment of policy goals. In order to achieve such an objective in a (semi-)automated way, the KP must have a knowledge representation for relating the different information artifacts (dataset, algorithms, etc.). It constitutes the ERA-PLANET abstract ontology which defines the major concepts and their relations.

Figure 10 depicts the Data to Knowledge for Policy (D2K4P) ontology divided into two packages: Abstract (including general concepts) and Processing (including concepts relevant for processing and Indicators generation). It includes the following concepts:

- The main concepts (in green):
    - **Observable**: a physical parameter which can be directly observed with proper instruments
    - **Essential Variable**: a physical parameter which is necessary to describe the Earth system status
    - **Indicator**: a derived parameter summarizing the status of the system under consideration. (Note that we use the term Indicator in a broad sense to include both an Indicator in strict sense – a physical parameter indicating the status of a system for decision-making purposes – and an Index – a figure summarizing multiple parameters to represent the status of a system for decision-making purposes).
    - **Policy Goal**: the desired outcome or what is to be achieved by implementing a policy
- The abstract connectors (in light blue):
    - **EV Generation Model**: a science-based process to generate Essential Variables from Observables. It includes both physical and statistical (including machine learning and AI) models since both have scientific foundness if correctly adopted.

    - **Indicator Generation Model**: a process to summarize Essential Variables representing the system status as an Indicator.
- The implementing concepts (in yellow):
    - **Dataset**: the collection of values of a specific Observable in the defined context (space, time, etc.)
    - **EV Value**: the values of the Essential Variable in the defined context
    - **Context Status**: the value of the Indicator summarizing the status of the defined context.
- The implementing connectors (in dark blue):
    - **EV Generation Algorithm**: an algorithm implementing the EV Generation Model
    - **Indicator Generation Algorithm**: an algorithm implementing the Indicator Generation Model

The general ontology allows the organization of knowledge handled in the KP. For example, a model stored in the KP is an instance of EV Generation Model, and it should be linked with the source code implementing the corresponding EV Generation Algorithm. Other links should provide information on the relevant instances of Observables and their available implementations as datasets. Therefore, collecting the information (models, datasets, etc.) and organizing it according to the general ontology allows reasoning (e.g. "which datasets can be used to run this model?", "Is it possible to generate this indicator based on these datasets?").

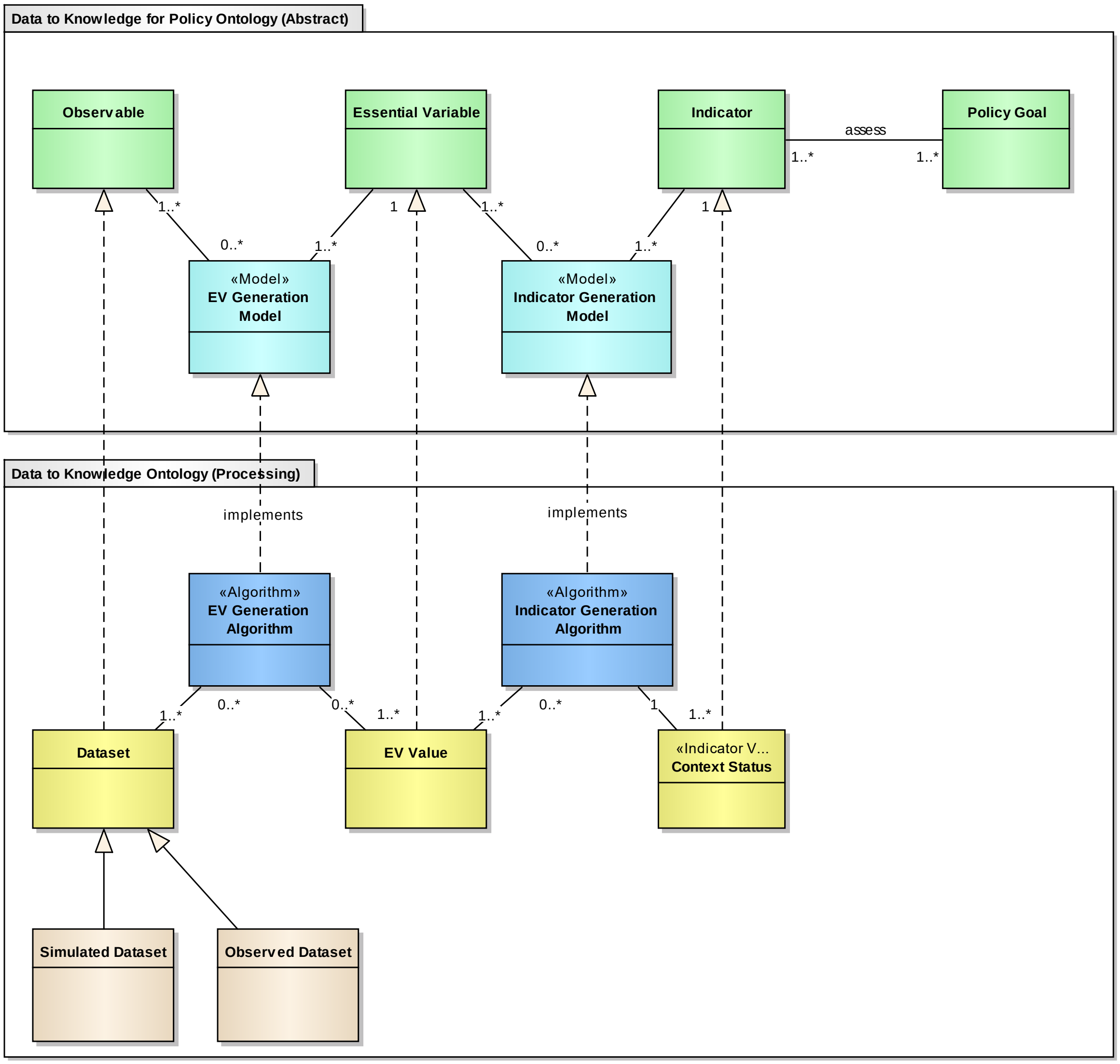

Figure 10 The Data to Knowledge for Policy (D2K4P) ontology

The Data to Knowledge for Policy (D2K4P) ontology describes the general process. It must be tailored to specific domains. For example, in iCUPE the Policy Goals may include the UN SDGs, the Paris Agreement on Climate Change objectives, etc., the relevant Essential Variables may concern the Biodiversity and Ecosystem, Weather and Climate, Ocean, etc. domains.

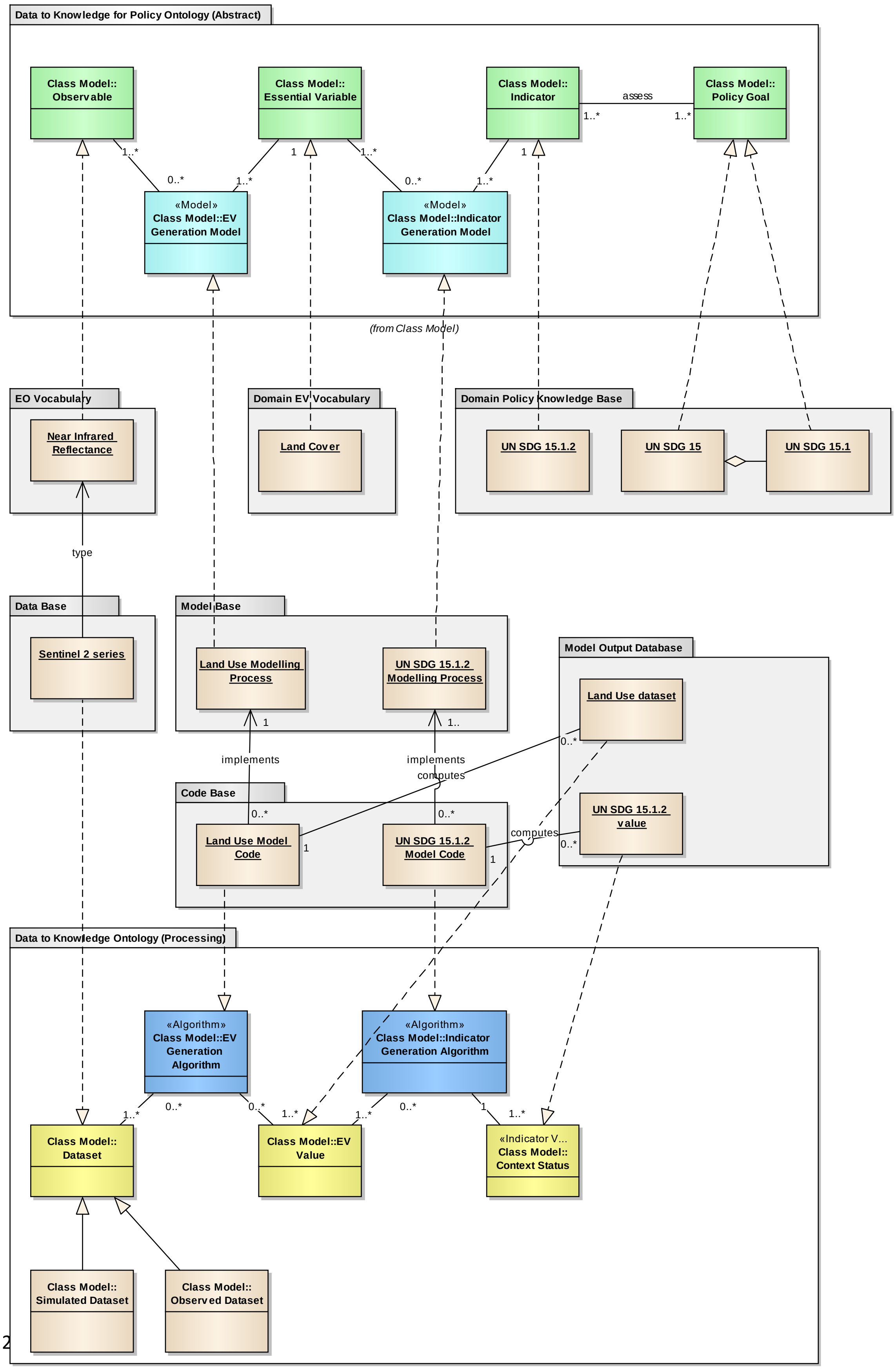
Data to Knowledge for Policy Ontology (Abstract)
Class Model:: Observable
Class Model:: Essential Variable
Class Model:: Indicator
Class Model:: Policy Goal
assess
«Model» Class Model::EV Generation Model
«Model» Class Model::Indicator Generation Model
(from Class Model)
EO Vocabulary
Near Infrared Reflectance
Domain EV Vocabulary
Land Cover
Domain Policy Knowledge Base
UN SDG 15.1.2
UN SDG 15
UN SDG 15.1
type
Data Base
Sentinel 2 series
Model Base
Land Use Modelling Process
UN SDG 15.1.2 Modelling Process
Model Output Database
Land Use dataset
UN SDG 15.1.2 value
implements
computes
Code Base
Land Use Model Code
UN SDG 15.1.2 Model Code
Data to Knowledge Ontology (Processing)
«Algorithm» Class Model::EV Generation Algorithm
«Algorithm» Class Model::Indicator Generation Algorithm
Class Model:: Dataset
Class Model::EV Value
«Indicator V... Class Model:: Context Status
Class Model:: Simulated Dataset
Class Model:: Observed Dataset

Figure 11 The ERA-PLANET Knowledge Base structure (with example packages and components)

Figure 11 shows an example of implementation of the internal ERA-PLANET Knowledge Base. It includes:

- **D2K4P Ontology (Abstract)**: the set of concepts and relations needed to model the Data to Knowledge transition
- **EO Vocabulary**: A vocabulary of parameters that can be considered as satellite, airborne, in-situ, etc. Observables (e.g. "Near Infrared Reflectance"). An example source is the OGC® "Earth Observation Metadata profile of Observations & Measurements" [29]
- **Domain EV Vocabularies**: Vocabularies of parameters that can be considered as Essential Variables for a specific domain. Example sources can be found from the activities of previous projects (e.g. ConnectinGEO), existing netwoks (e.g. ENEON), and parallel activities in ERA-PLANET.
- **Domain Policy Knowledge Base**: Knowledge bases storing the goals, targets, and indicators defined for a specific domain (e.g. the UN SDGs).
- **Data Base**: Databases of datasets
- **Model Base**: A knowledge base storing information on available modelling processes (flow diagrams, algorithms) to generate Essential Variables and Indicators.
- **Code Base**: databases storing the information (e.g. service endpoints, source code) about the executable processes implementing Models.
- **Model Output Database**: a database storing the datasets generated by Algorithm runs.

The implementation of the ERA-PLANET Knowledge Base consists in collecting the relevant packages and establish the necessary links to encode the previously tacit knowledge (i.e. alignment of Vocabularies with the D2K4P Ontology, association of objects to the correct concept).

For example, each time a new model is added to the Model Base, it must be specified whether it is a realization of a EV Generation to generate an Essential Variable or a realization of a Indicator Generation Model to generate an Indicator. Moreover, also the conceptual links between abstract concepts must be realized. For example, if the model is EV Generation, it must be specified wich Observables it needs, and which Essential Variables it generates.

5.5 Engineering Viewpoint

Engineering VP is concerned with the infrastructure required to support system distribution.
[25]

Figure 12 shows the engineering view of the ERA-PLANET architecture.

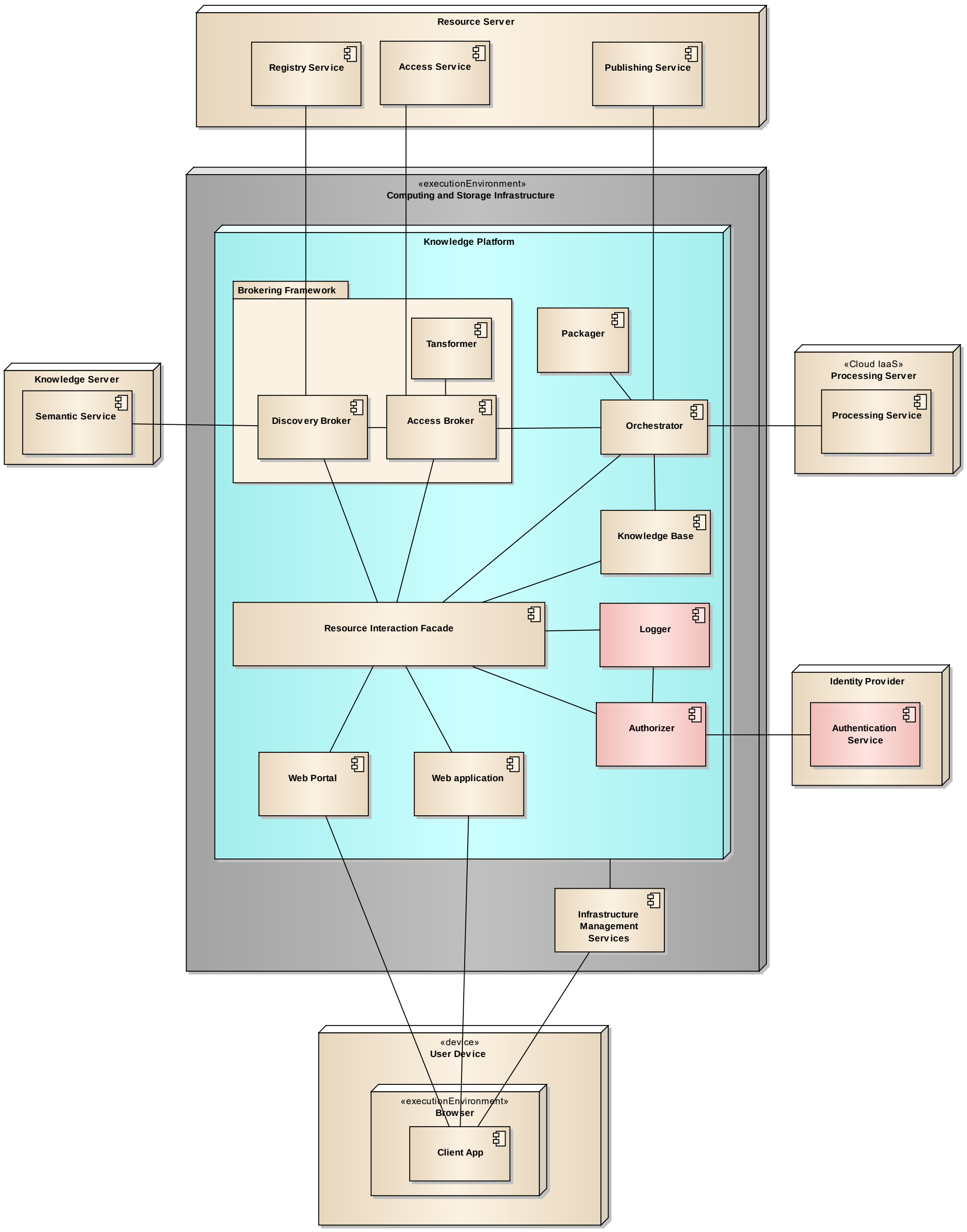

Figure 12 Engineering view of the ERA-PLANET Knowledge Platform architecture

The iCUPE architecture includes a set of different nodes:

Node	Description
Resource Server	A Resource Server is a node dedicated to serve resources of different types (datasets, algorithms, etc.). ERA-PLANET 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
Knowledge Platform	The Knowledge Platform is the core architectural node. It contains all the components and tools needed to achieve the iCUPE objective
Computing and Storage Infrastructure	A Computing and Storage Infrastructure is a node hosting the Knowledge Platform. It may be either a node managed locally by one ERA-PLANET partner, or a private or public cloud offering Infrastructure-as-a-Service capabilities
Knowledge Server	A Knowledge Server is a node providing services accessible by the Knowledge Platform for semantic discovery 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
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, algorithm source code, etc.). They range from plain download services to full access services supporting data 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, algorithm source code, etc.)
Processing Service	Processing Services enable the execution of processing algorithms and workflows, generating products that are stored as new resources
Brokering Framework	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. • A Transformer 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.)
Semantic Service	Semantic Services expose knowledge-bases such as thesauri, gazetteers, ontologies, allowing to find terms related to a keyword for query expansion

Packager	The Packager builds a software container for the algorithm implementing a model
Knowledge Base	The Knowledge Base hosts the knowledge about available models and datasets according to the ERA-PLANET ontology
Orchestrator	The Orchestrator is the core component for knowledge generation. Basing on the Knowledge Base content, it implements the general workflow: a) retrieve the algorithm source code from the Resource Server, b) call the Packager to build the software container, c) send the container to a Remote Processing Service; d) collect the necessary input datasets from the Resource Servers; e) invoke the Remote processing Service, f) collect the output, g) publish the result on a Resource Server
Resource Interaction Facade	The Resource Interaction Facade is a component that aggregates the different components exposing a simplified interface facilitating the interaction with the KP
Infrastructure Management Services	The Infrastructure Management Services are provided by the Computing and Storage Infrastructure. They allow a KP Administrator to manage the KP 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. Different portals serve different user categories (the Dashboard for policy-makers, the Development Portal for modelists, etc.)
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 components behind the Facade
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 KP Platform

5.5.1 The knowledge generation process in the ERA-PLANET KP

Figure 13 shows the activity diagram for the knowledge generation processing. Basically, when a user, through the GUI asks running a model, the Knowledge Platform retrieves the necessary source code, and creates a package including the executable code and the necessary library. Then, it sends it to a Remote Processing Server which exposes it as a remote processing service. The Knowledge Platform calls for the service, which generates the output. Through the GUI it is possible to visualize the output.

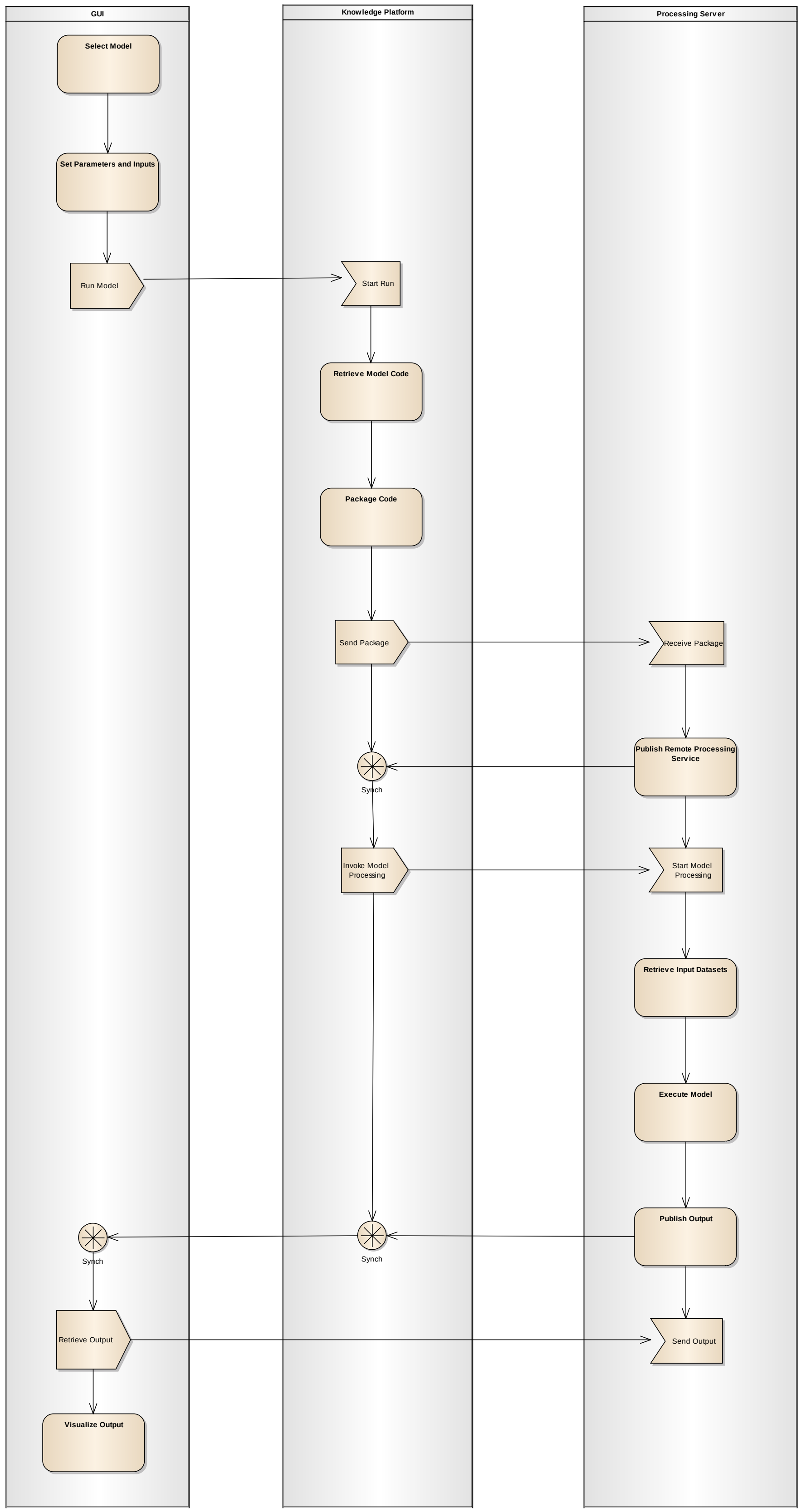
GUI
Knowledge Platform
Processing Server
Select Model
Set Parameters and Inputs
Run Model
Start Run
Retrieve Model Code
Package Code
Send Package
Receive Package
Publish Remote Processing Service
Synch
Invoke Model Processing
Start Model Processing
Retrieve Input Datasets
Execute Model
Publish Output
Synch
Synch
Retrieve Output
Send Output
Visualize Output

Figure 13 Activity diagram for the generation of knowledge (only major components and activities shown)

5.5.2 The iCUPE Knowledge Platform in the ERA-PLANET context

iCUPE is an ERA-PLANET Transnational Project. The other three strands have similar activities to design and develop platform for data, information and knowledge generation and sharing. Although the requirements differ, synergy between the efforts in Transnational Project is useful and even necessary, for avoiding duplication of efforts, optimizing mobilization of resources and last but not least, a better and harmonized contribution of ERA-PLANET to international inititatives – GEO/GEOSS in primis.

Taking into account the preliminary design of platforms in the other three Transnational Project, a common design principle is the adoption of a loosely-coupled architecture allowing interoperability with the external environment - see ERA-PLANET DP6 "ERA-PLANET Knowledge Platform exposes a set of (high-level) APIs for interaction with the external environment". Therefore, in general it is possible to conceive an interaction between the different Transnational Project platforms.

Moreover, the different platforms share a second design principle concerning the extension of the platform itself through internal APIs – see ERA-PLANET DP4 "ERA-PLANET Knowledge Platform is made of software components interacting through (low-level) APIs". Therefore, it is possible to conceive the four platforms as specialization through extension of a single ERA-PLANET Platform.

Thus, two general solution are possible in the ERA-PLANET context:

- A single ERA-PLANET Knowledge Platform with dedicated views and functionalities for the four Transnational Projects.
- Four Transnational Project platforms interoperable through APIs.

However, considerations about simplicity of management and scalability suggest to adopt the first approach with a single ERA-PLANET Knowledge Platform with dedicated views, and personalization for the four Transnational Projects as shown in Figure 14.

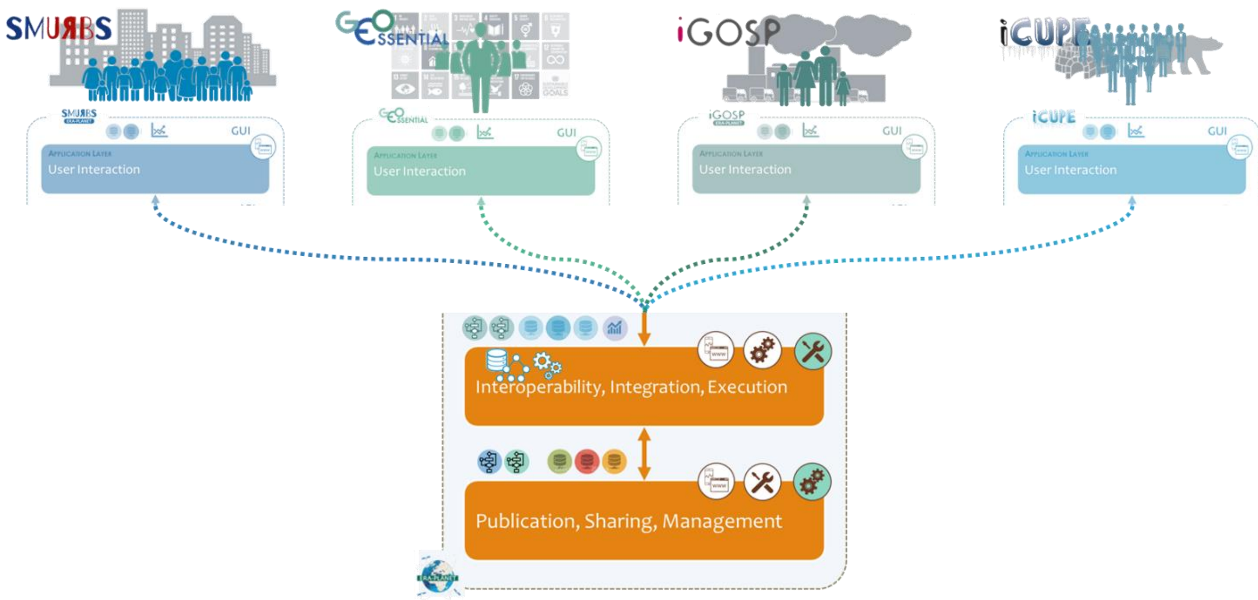

Figure 14 Scenario 1: single ERA-PLANET KP instance with multiple views for Transnational Projects

Figure 15 shows the current implementation schema of the ERA-PLANET Knowledge Base relating the basic D2K4P ontology and the internal resource models of the ERA-PLANET KP.

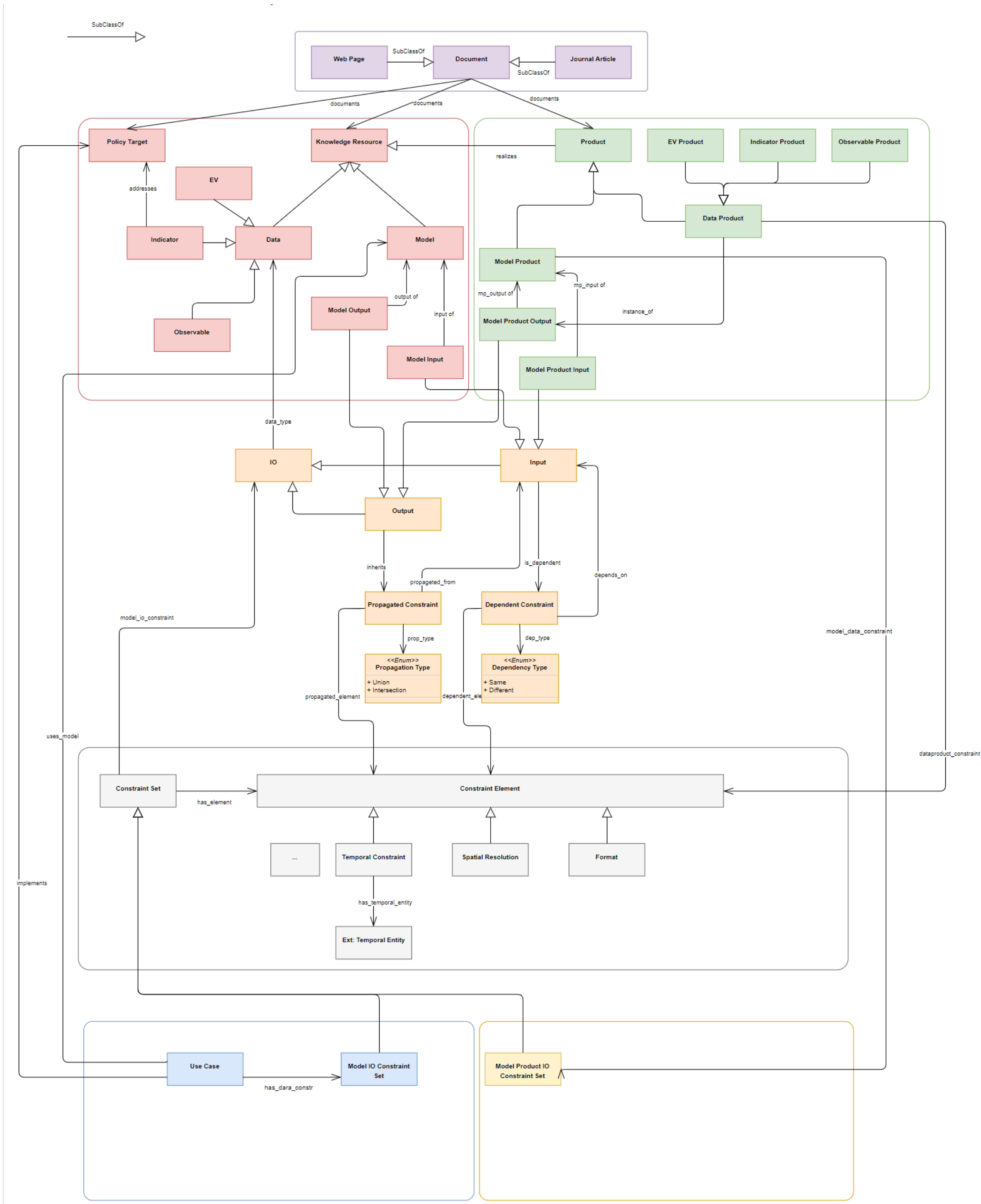

Figure 15 Logical schema of the ERA-PLANET Knowledge Base. The diagram relates the basic Data to Knowledge for Policy ontology (in red) with the internal VLab resource models.

5.5.3 Interoperability with GEOSS

ERA-PLANET has a specific requirement concerning GEOSS. EDMP-3 reads: "All data generated in the action, which are relevant, directly or indirectly, for information to policy and decision-makers in key societal benefit areas must be accessible through GEOSS and Copernicus at the conditions described in the DMP and in compliance with GEOSS-DSP and GEOSS-DMP". This translates in the FR12 specific requirement: "Data registration in GEOSS". It implies that the ERA-PLANET KP (or individually the GEO-ESSENTIAL KP) must be registered as a GEOSS Data Provider, so that every dataset generated in the KP and stored, locally or remotely, can be published and made visible through GEOSS. (By a technical point-of-view the ERA-PLANET KP can support different workflows for the publication in GEOSS of model outputs; the specific process adopted in SMURBS will be defined by the SMURBS Consortium.)

However, GEOSS is also one of the relevant data sources for the ERA-PLANET KP, therefore the communication between GEOSS and the ERA-PLANET KP should be a two-way interaction.

Leveraging the experience of interoperability between the ECOPOTENTIAL VLAB and the GEOSS Platform, Figure 16, shows the proposed approach for interoperability between the ERA-PLANET KP and GEOSS:

- The GEOSS Platform exposes discovery and access capabilities through the GEOSS APIs. The ERA-PLANET KP can use the GEOSS APIs to make GEOSS resources available in the KP.
- The ERA-PLANET KP exposes its discovery, access and knowledge generation capabilities through the ERA-PLANET KP APIs. The GEOSS Platform can use these APIs to retrieve resources and, in general, content useful to build Community Portals, or SBA Knowledge Bodies. Of course, this step requires to interact with the GEOSS Platform technical team, but this is feasible since ERA-PLANET is one of the initiative selected by the H2020 EDGE project aiming at operating and enhancing the GEOSS Platform core components: the GEOSS Portal and the GEO DAB.

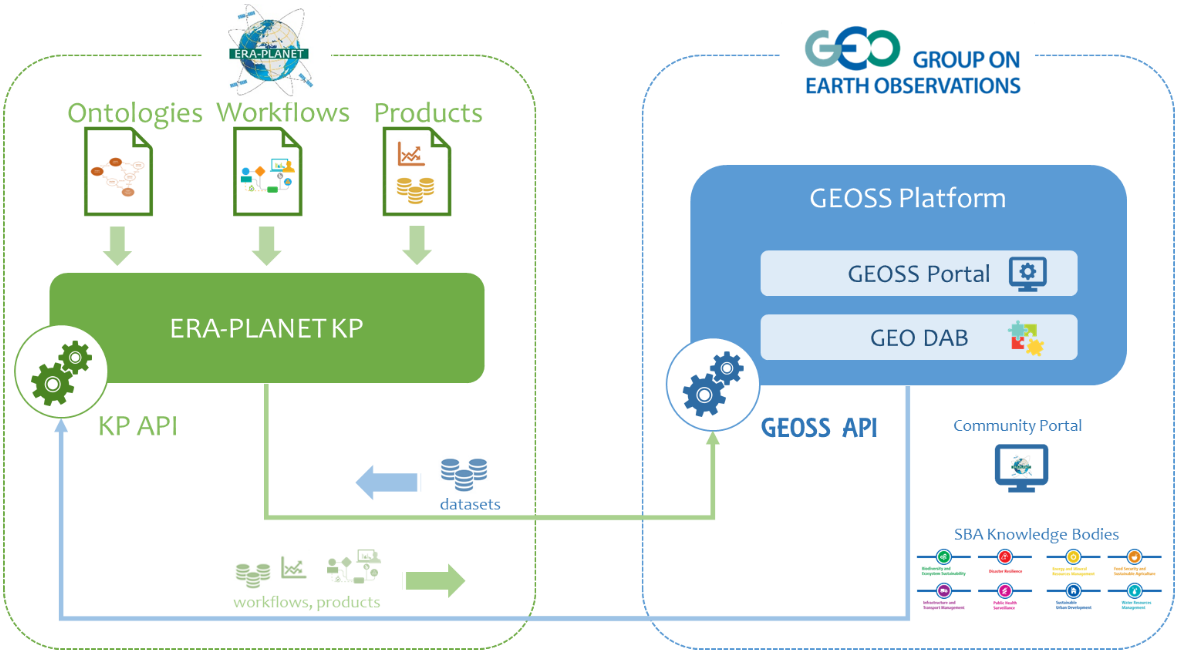

Figure 16 Interoperability between the ERA-PLANET KP and the GEOSS Platform

5.6 Technology Viewpoint

Technology VP is concerned with the choice of technology to support system distribution.
[25]

The ERA-PLANET KP will be implemented using and extending existing solutions and tools. At the time of the ERA-PLANET proposal and DoW preparation some key technologies were identified as ERA-PLANET Key Enabling Technologies. Other technologies, developed in the context of previous research and innovation initiatives and projects, are provided or under control of iCUPE partners and will be reused in iCUPE.

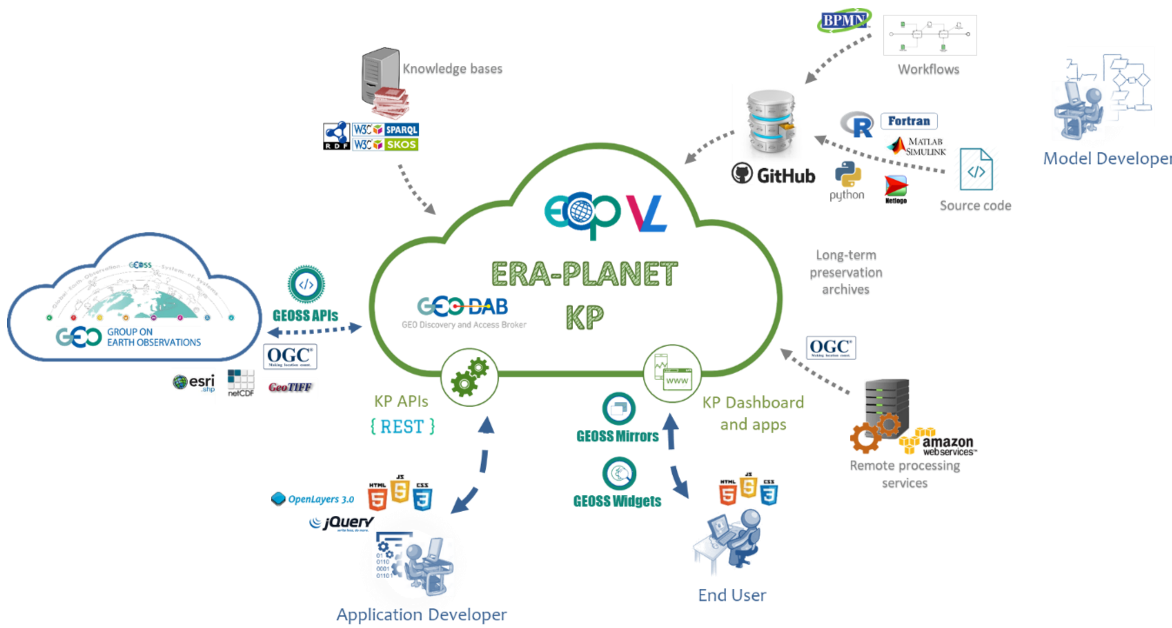

Figure 17 ERA-PLANET Knowledge Platform technologies (not exhaustive)

5.6.1 ERA-PLANET Key Enabling Technologies and horizontal actions

In keeping with the ERA-PLANET interoperability principles, iCUPE will implement and

make use of the following Key Enabling Technologies (KETs) with the objectives of:

1 **Applying the GCI brokering approach and its principles to implement multidisciplinary interoperability** and lower entry barriers for both Users and Data Providers. iCUPE will promote GEOSS Data Management and Data Sharing Principles and make use of the GEO Discovery and Access Broker (DAB) technology together with its Application Programming Interfaces (APIs) (Santoro et al. 2016).

2 **Specifying and enhancing shared data & service quality**. This means considering Quality Assurance (QA), Quality Control (QC) and Quality of Services (QoS) exemplified by the work endorsed by CEOS and GEO on Quality assurance framework for earth observation (QA4EO) (http://qa4eo.org). The project will also consider other aspects such as fit for purpose indicators using GeoViQua models and QualityML; annotations and geospatial user feedback capabilities; Persistent Identifiers (PID) and data citation.

3 **Including re(use) metadata for shared data & services**. The project is aiming at using and reusing what is existing and available through metadata description. Semantic description and "how" metadata in addition to "what" ones can leverage the use of existing resources.

4 Technical and semantic interoperability are essential to facilitate data discovery, access and integration. However, another important aspect of interoperability concern issues related to intellectual property of data, in particular for reducing data fragmentation. The Research Data Alliance (RDA) and the International Council for Science: Committee on Data for Science and Technology (ICSU-CODATA) have recently published a report on Legal Interoperability Of Research Data: Principles and Implementation Guidelines (October 2016 - https://doi.org/10.5281/zenodo.162241) that will be of interest for iCUPE.

5 **Data, services, and information generated in the frame of the project will directly contribute to GEOSS via the GCI** using as much as possible the GEO DAB APIs and the ENEON commons.

6 ERA-PLANET will pursue synergies with **Copernicus DIAS**, Copernicus services and ESA TEPs.

7 Finally, the project with benefit and contribute to the European Network of Earth Observation Networks (ENEON) to coordinate in situ EO networks and facilitate the use of data sharing principles across networks.

In **bold**, the most relevant points by a technology point-of-view are highlighted.

Table 7 sumarizes the technological choices for the implementation of the ERA-PLANET KP. The following paragraph detail and motivate the most relevant choices.

Table 7 - ERA-PLANET KP technological choices

Component	Selected Technology
Registry Service	Registry Services are external services. The Brokering Framework supports the most common de iure and de facto specifications for inventories, catalogues, and metadata profiles/format. The Brokering Framework is also extensible to specifications not supported yet
Access Service	Access Services are external services. The Brokering Framework supports the most common de iure and de facto specifications for access and visualization services. The Brokering Framework is also extensible to specifications not supported yet
Publishing Service	iCUPE will implement/select a Resource Server for publishing processed data. The Resource Server will be connected to the ERA-PLANET KP for discovery and access, possibly through connection with GEOSS. iCUPE will implement the Publishing Service selecting the most suitable specification
Processing Service	Processing Services are external services. The ERA-PLANET KP will support at least Copernicus DIAS, European Open Science Cloud (EOSC), Amazon WS for IaaS, and it will investigate the support of other technologies as outcomes of other on-going projects (e.g. H2020 EOSC-hub for porting on the EOSC)

Brokering Framework	The ERA-PLANET KP will run a dedicated instance of the GEO DAB Discovery and Access protocol
Semantic Service	The ERA-PLANET KP will support external Semantic Services exposing SPARQL/SKOS interface
Packager	The Packager will support at least the Docker container technology. iCUPE will investigate the support of other technologies.
Knowledge Base	The Knowledge Base will be based on OWL/RDF technologies. It will be implemented on a triple store based on technology to be identified
Orchestrator	iCUPE will build upon the ECOPOTENTIAL Virtual Laboratory solution and technology
Resource Interaction Facade	The Resource Interaction Facade is implemented exposing: a) standard service interface offered by the Brokering Framework; b) RESTful APIs; c) Web APIs (HTML5+Javascript+CSS library). The GEO DAB already supports the most common de iure and de facto standard service interfaces (including OGC/ISO). The APIs will extend those already exposed/offered by the GEO DAB and the ECOPOTENTIAL VLab
Infrastructure Management Services	The ERA-PLANET KP will be deployed first on Amazon IaaS, and it will use the management services offered by Amazon. ERA-PLANET will also investigate the support of other technologies (Copernicus DIAS, H2020 EOSC-hub for porting on the European Open Science Cloud)
Web Portal	iCUPE will use initially the general community portal based on the GEOSS Platform technologies (GEOSS Mirror Portal) developed in the context of the H2020 EDGE project.
Web Application	Dedicated Web Applications can be built on top of the ERA-PLANET KP based on the Web APIs
Browser	The Web APIs will support the major browsers.

Authentication Service	The ERA-PLANET KP will support multiple authentication providers including the most widespread (Google, Facebook) and other to be identified
Authorizer	The Authorizer will be implemented internally to support different profiles
Logger	The Logger will aggregate information fromthe existing logging services for the different components, including the Amazon AWS, internal GEO DAB insance, internal Web server, etc.

5.6.2 GEO Discovery and Access Broker (DAB) Framework

The **GEO Discovery and Access Broker** (GEO DAB) is based on the GI-suite Brokering Framework, 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 iCUPE 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 iCUPE.)

The choice of the GI-suite Brokering Framework as the central brokering component of the KPs 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 iCUPE 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 iCUPE consortium.

5.6.3 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:

"GEMET - Concepts, version 2.4"
"GEMET - Groups, version 2.4"
"GEMET - INSPIRE themes, version 1.0"

"GEMET - Supergroups, version 2.4"
"GEMET - Themes, version 2.4"
"GEOSS - Earth Observation Vocabulary, version 1.0"
"GEOSS - Societal Benefit Areas, version 1.0"
"INSPIRE - Feature Concept Dictionary, version 3"
"INSPIRE - Glossary, version 3"
"ISO - 19119 geographic services taxonomy"

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.

5.6.4 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 iCUPE 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 iCUPE, 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 KP 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 KP 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 KP 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

The iCUPE 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 iCUPE. In particular:

- For *datasets*, the major de iure and de facto standard specifications for service interfaces and metadata/data model/encoding are supported.
- For *models*, their representations will be stored in Business Process Model and Notation (BPMN) with augmented metadata according to the ECOPOTENTIAL VLab conventions;
- For *algorithms*, the Git specifications are supported to access source code from Git repositories, and OGC WPS is supported to access online processing services.

5.6.7 Packager

The Packager will be based on the Docker packaging component of the original ECOPOTENTIAL VLab.

5.6.8 Knowledge Base

The Knowledge Base will be implemented using a triple store and RDF reasoner, that is a database storing RDF triples representing the concepts and relations of the required concepts and relations. RDF triples will be harvested from existing databases in different formats (relational databases, tables, etc.) where already existing. For those relations that are not necessary for reasoning, they can be implemented tagging the resource metadata. For example, to associate an Essential Variable type to a dataset the already harvested, dataset metadata can be enriched, instead of creating a specific RDF triple.

5.6.9 Orchestrator

The Orchestrator will be based on the orchestrating component of the ECOPOTENTIAL VLab.

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 18) [30] 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) [31]:

- 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, iCUPE has great requirements at least on privileging "working software", "customer collaboration" and fast "response to change". Therefore, iCUPE will adopt an Agile Methodology for design and development.

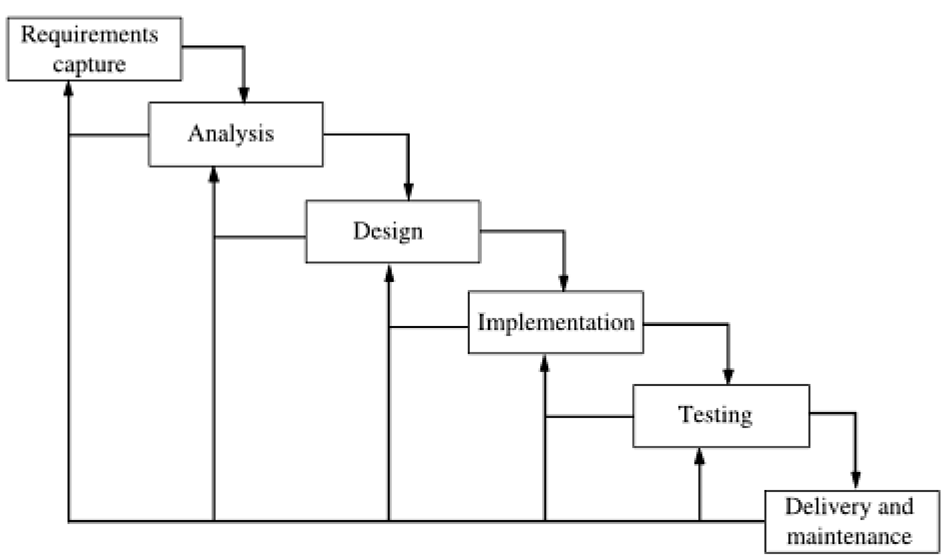

Figure 18 The traditional Waterfall Model (from [30])

Agile methodologies better respond to changes through an iterative process (Figure 19). 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.

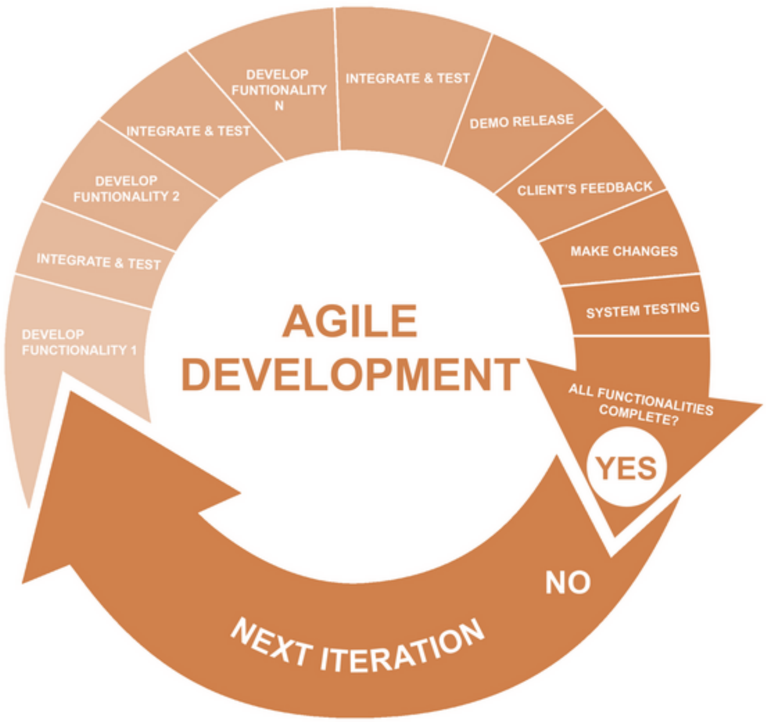

Figure 19 The iterative process in Agile development

Taking into account the specificity of ERA-PLANET we can identify two main milestones and therefore two main iterations:

- Project Month 18, end of the first major iteration and release of first version of the ERA-PLANET KP.
- Project Month 30, end of the second major iteration and release of the final version of the ERA-PLANET KP.

Other milestones can be defined to address specific events (e.g. GEO Weeks).

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 KP capacity

6.2 System integration

As described in Section 4.1, ERA-PLANET Knowledge Platform 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 Knowledge Platform architecture are then implemented through the integration of selected technological solutions to build a complete framework delivering the requested Knowledge Platform functionalities. The first release of the Knowledge Platform comprised the GI-suite Brokering Framework. The following releases will include selected components integrated with the GI-suite Brokering Framework to support missing functionalities.

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 iCUPE are:

- *Discovery broker (GI-cat)*: a component which can 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 can 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 20):

- The *Distributor* oversees 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.

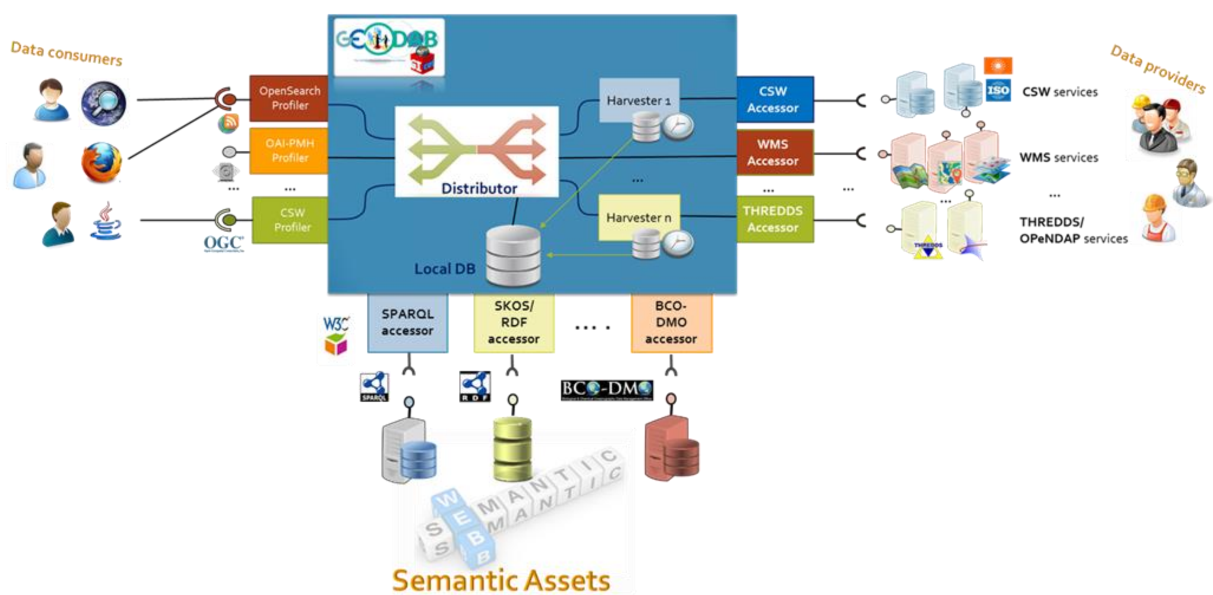

Figure 20 Discovery broker (GI-cat) internal architecture

- *Semantic Enhancement Module (GI-sem)*: a component which implements semantic query expansion [28]. 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 can 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-axe includes several modules (see Figure 21):
    - The *Orchestrator* oversees 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 considering 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.

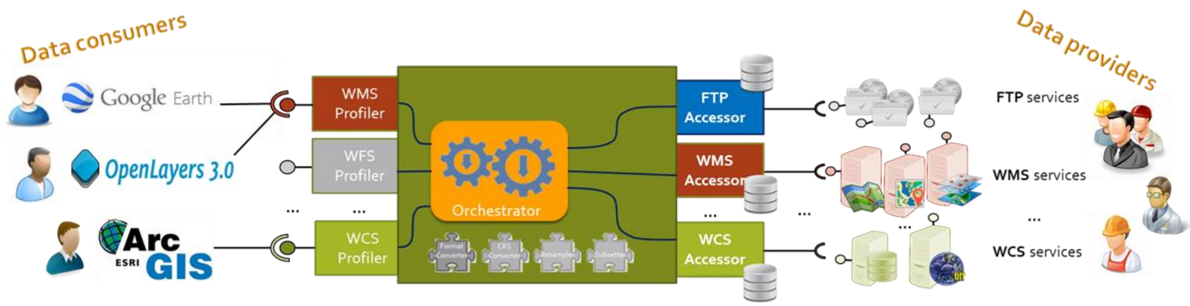

Figure 21 Access Broker (GI-axe) internal architecture

- *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.

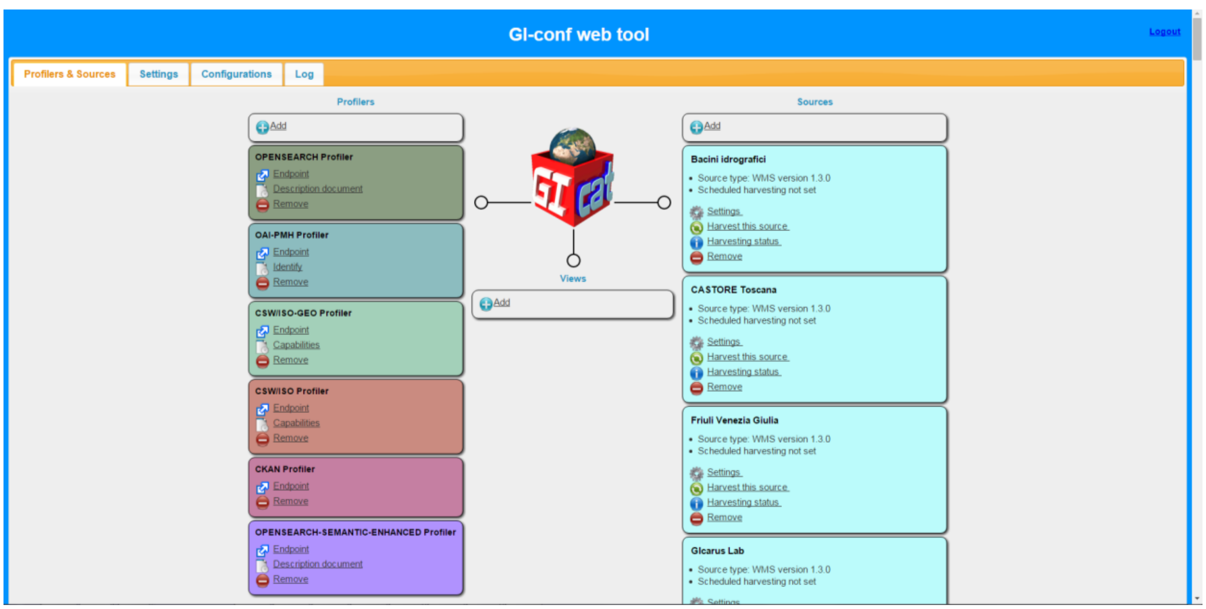

Figure 22 GI-conf screenshot

- *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 CSW 2.0.2 ebRIM/CIM	Discovery interface
GeoNetwork (versions 2.2.0 and 2.4.1) catalog service	Discovery interface
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
THREDDS-NCISO 1.0.1, 1.0.2	Discovery and access interfaces, and metadata model
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 Global Mercury Observation System	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	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	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
Gestione Integrata e Interoperativa dei Dati Ambientali	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
GEOSS EuroGEOSS A EUROPEAN APPROACH TO GEOSS	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
	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 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
	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 Knowledge Platform). It also exposes APIs for configuration and notification. The GI APIs facilitate the use of discovery and access functionalities by intermediate users (developers).

6.3 Status of implementation

The status of implementation at the iCUPE project-month 10 (January 2020) is summarized in the following tables listing the major components:

Node	Status
Resource Server	[On-going analysis] A major (logical) Resource Server is GEOSS. Other Resource Servers will be identified in the preparation of the DMP
Knowledge Platform	[Under test] The Knowledge Platform is deployed on Amazon Virtual Machines
Computing and Storage Infrastructure	[Under test] The selected Computing and Storage Infrastructure are Copernicus DIAS, EOSC, Amazon and UNIGE cluster Baobab. Copernicus DIAS, EOSC and Amazon already tested and working.

Knowledge Server	[Under evaluation]. The selected Knowledge Server is the EC-JRC Semantic Service, but is discontinued
User Device	[Mature] Common desktop and mobile devices are supported
Identity Provider	[Under test in GEOSS] Common Identity Providers are supported by the selected frameworks (e.g. Google, Facebook, OpenAuth)

Component	Description
Registry Service	[On-going] The DMP will inform about the adopted specifications. The Brokering Framework already support the most widespread specifications for registry services.
Access Service	[On-going] The DMP will inform about the adopted specifications. The Brokering Framework already support the most widespread specifications for access services.
Publishing Service	[On-going] The DMP will inform about the adopted specifications
Processing Service	[Under test] The OGC WPS specification has been selected, and already supported by the Orchestrator
Brokering Framework	[Operational] The selected technology is the same of the GEO DAB
Semantic Service	[Operational] The Brokering Framework already supports query expansion from knowledge bases exposed as SPARQL/SKOS
Packager	[Operational] The selected Packager component already supports Docker packaging
Knowledge Base	[On-going analysis] A triple-store must be selected. Ontologies must be defined and knowledge base encoded
Orchestrator	[Operational; enhancement on-going] The selected Orchestrator supports the invocation of the Packager, and remote services. It must be enhanced with the integration with the Knowledge Base

Resource Interaction Facade	[Operational; enhancement on-going] The Resource Interaction Facade is based on GEOSS APIs, and on-going activities on the ECOPOTENTIAL Knolwedge APIs.
Infrastructure Management Services	[Mature] Amazon management services already supported
Web Portal	[On-going] See activities on the iCUPE Dashboard, GEOSS Community Portal, dedicated applications for EuroGEOSS Sprint-to-Ministerial
Web Application	[Analysis] The design of the user interfaces (including the Dashboard) will define requirements for server-side applications
Browser	[Mature] Common browsers are supported

6.3.1 VLab General Portal

A VLab General Portal is under release. It aims at providing a single entry point for the three main stakeholder categories: VLab end users, VLab application developers, VLab modelers.

It worths noting that VLab is considered as the marketing name for the technology developed from the ECOPOTENTIAL VLab through the ERA-PLANET Knowledge Platform enhancements.

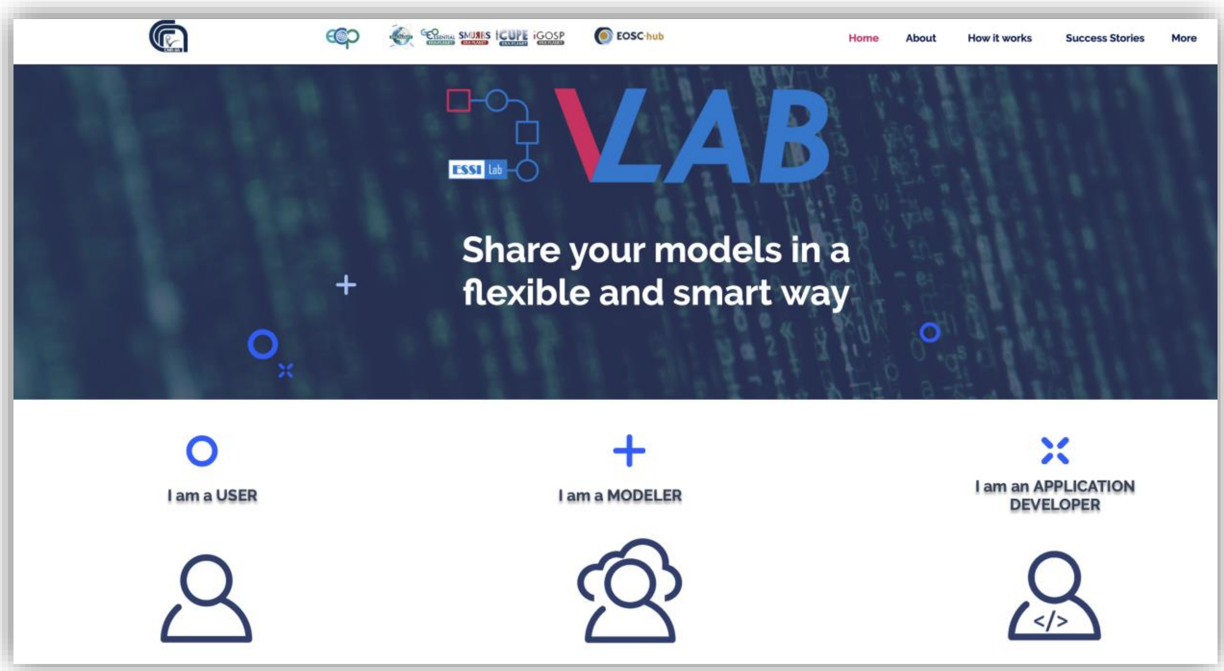

Figure 23 - Screenshot from the VLab Portal

6.3.1 VLab Portal for Modelers

The VLab Portal for modelers (https://vlab.geodab.org) has been re-designed and the documentation has been updated.

Figure 24 - Screenshot from the new VLab Portal for modelers

6.3.1 D2K4P Ontology

A first version of the Data-to-Knowledge-for-Policy ontology has been detailed based on a couple of use-cases. It is under refinement and it is described in the iCUPE D1.2 deliverable series.

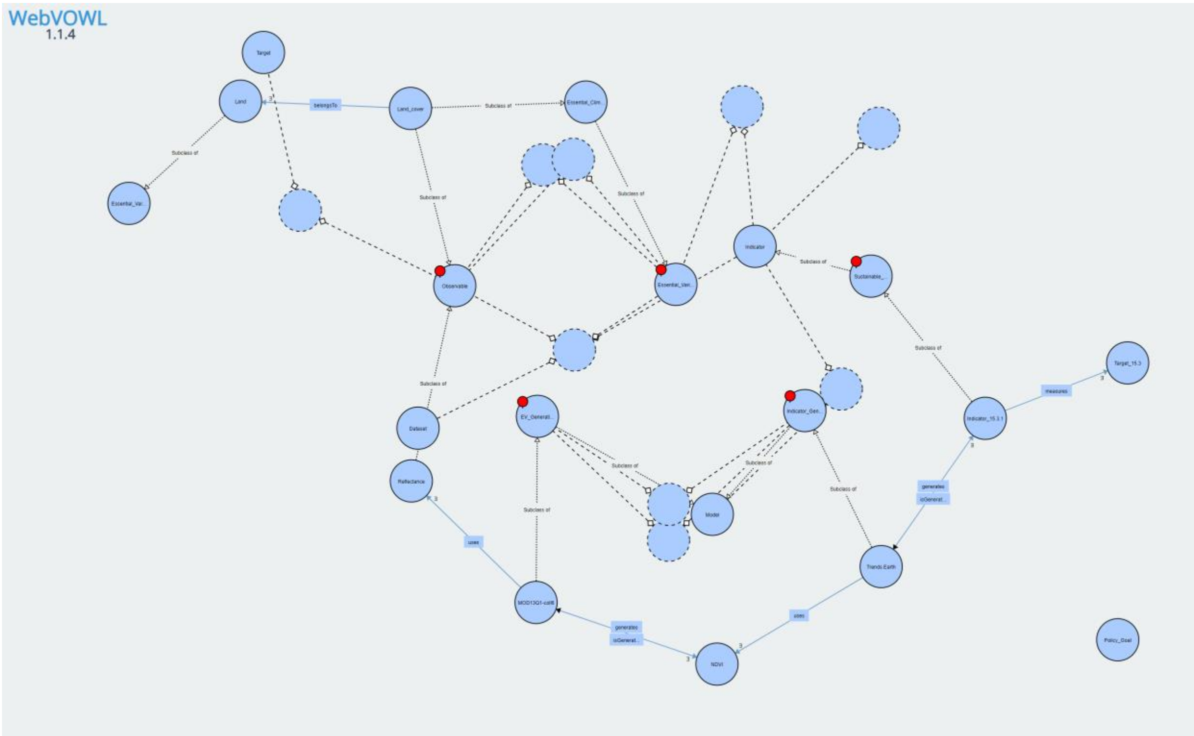

Figure 25 - The draft of the D2K4P ontology instance for the Land Degradation use-case

6.3.1 Model Porting

Currently, the VLab gives access to 21 public workflows developed in the context of different projects and initiatives. Other are still under test and not publicly available.

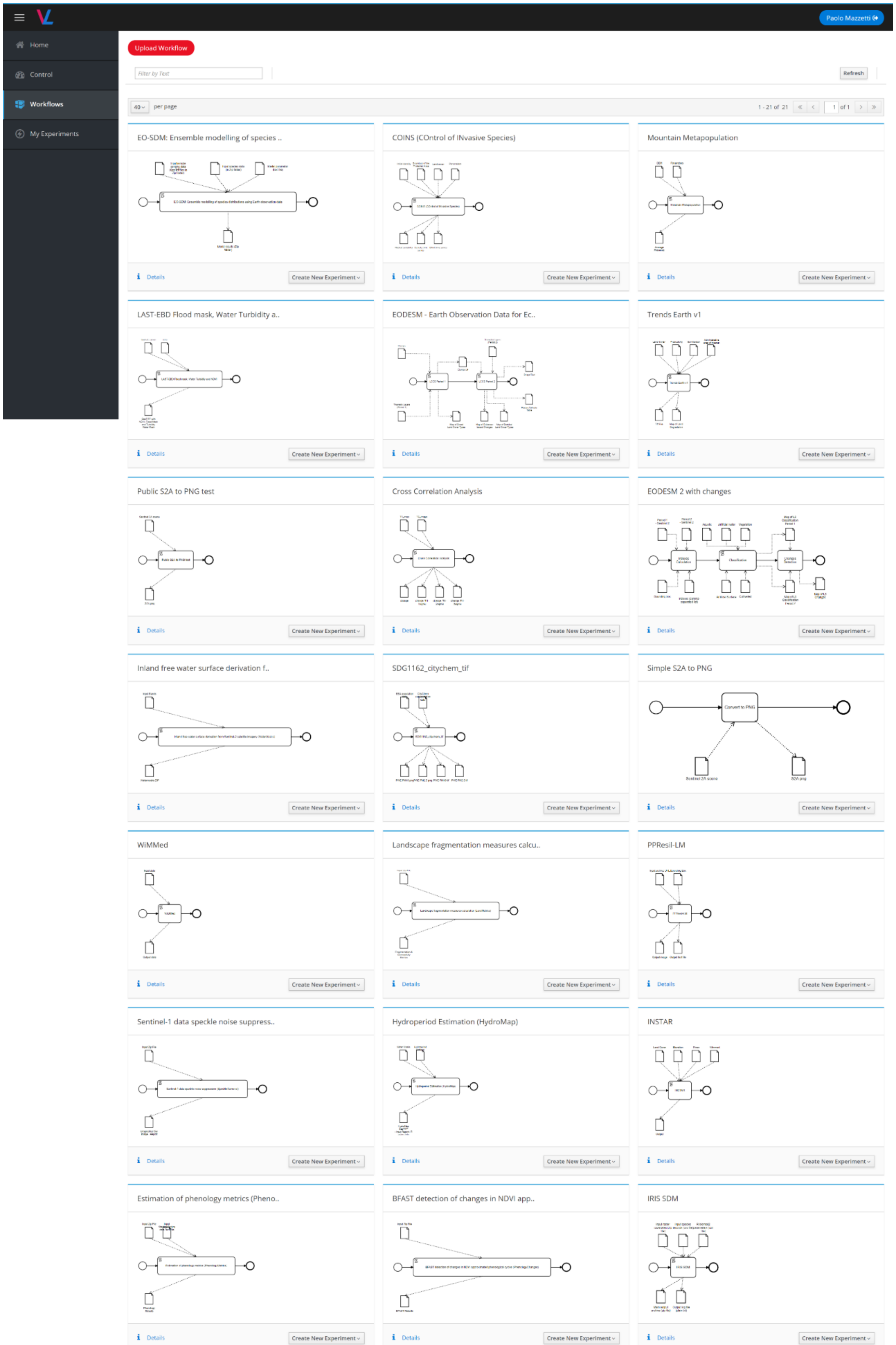
Paolo Mazzetti
Home
Control
Workflows
My Experiments
Upload Workflow
Filter by Text
Refresh
40
per page
1 - 21 of 21
EO-SDM: Ensemble modelling of species ..
COINS (COntrol of INvasive Species)
Mountain Metapopulation
LAST-EBD Flood mask, Water Turbidity a..
EODESM - Earth Observation Data for Ec..
Trends Earth v1
Public S2A to PNG test
Cross Correlation Analysis
EODESM 2 with changes
Inland free water surface derivation f..
SDG1162_citychem_tif
Simple S2A to PNG
Convert to PNG
WiMMed
Landscape fragmentation measures calcu..
PPResil-LM
Sentinel-1 data speckle noise suppress..
Hydroperiod Estimation (HydroMap)
INSTAR
Estimation of phenology metrics (Pheno..
BFAST detection of changes in NDVI app..
IRIS SDM
Details
Create New Experiment

Figure 26 - The workflows publicly accessible from the VLab Portal for modelers

6.3.2 VLab based applications

Three different applications have been developed/enhanced using the VLab APIs.

GEOEssential dashboard

The GEOEssential Dashboard is an example of dedicated application that allows for interactive exploration of the output of VLab model runs. In particular, it has been tailored for Land Degradation reporting using the Trends.Earth model porting available in the VLab.

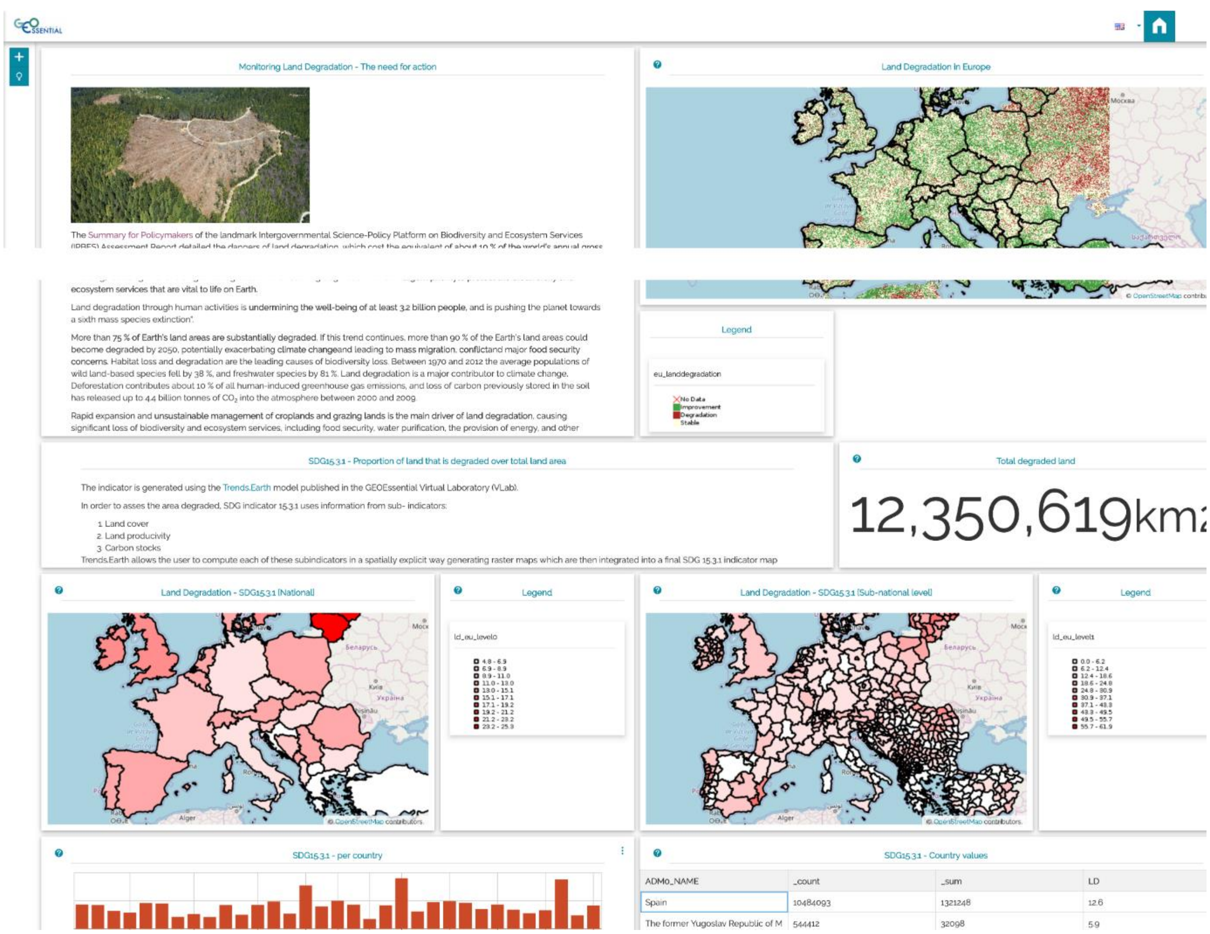

Figure 27 - Screenshot from Land Degradation in Europe report usig the iCUPE Dashboard

GEOSS ECOPOTENTIAL Community Portal

The GEOSS ECOPOTENTIAL Community Portal has been redesigned according to the Results-Oriented GEOSS vision, allowing running and access of the EODESM model for Land Cover and Land Cover Change.

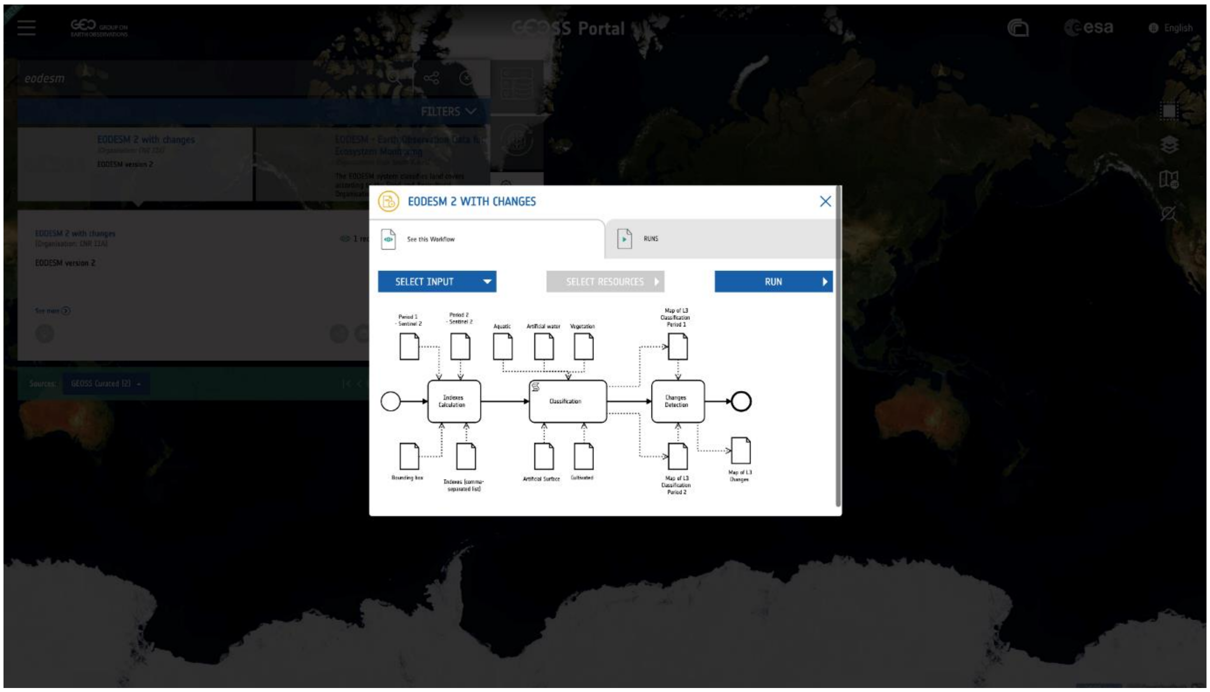

Figure 28 - Selection of the EODESM model ported on the VLab from the GEOSS ECOPOTENTIAL Community Portal

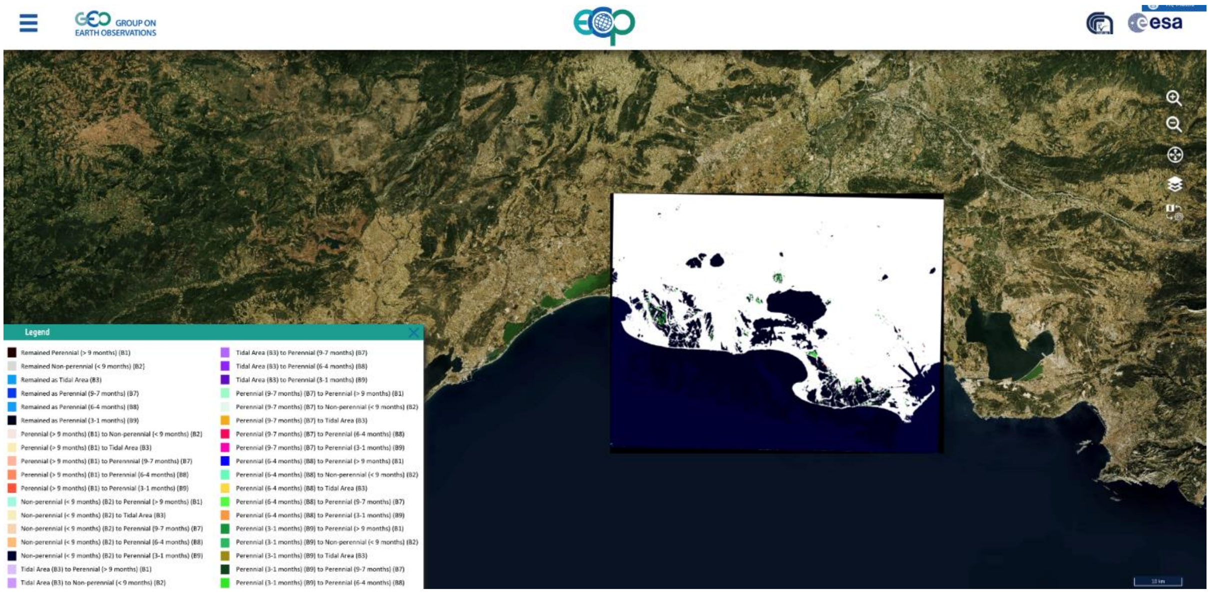

Figure 29 - Output of the EODESM model running through the VLab in the GEOSS ECOPOTENTIAL Community Portal

Protected Area Analysis application

In the context of the EuroGEOSS Sprint-to-Ministerial a dedicated application for running different models on different clouds (Copernicus DIAS, EOSC, Amazon) using the VLab has been developed as a joint effort of different organizations and initiatives.

Figure 30 _ Output of the EODESM Model running on a Copernicus DIAS using the VLab through the Protected Area Analysis application

7 References

[1] ECOPOTENTIAL Consortium, "Design of the ECOPOTENTIAL Virtual Laboratory - D10.1," 2016.

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

[3] European Commission, "TOPIC : Strengthening the European Research Area in the domain of Earth Observation," 2013. [Online]. Available: https://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/sc5-15-2015.html. [Accessed 24 05 2018].

[4] European Commission, "Grant Agreement Number - 689443 - ERA-PLANET," 2015.

[5] iCUPE Consortium, "Consortium Agreement for iCUPE Project," 2017.

[6] ERA-PLANET Consortium, "Data Management Plan - D4.5," 2018.

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

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

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

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

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

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

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

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

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

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

[26 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].

[27 iCUPE Consortium, "iCUPE data management plan (ver 2)," 2018.

[28 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.

[29 Open Geospatial Consortium, "OGC® Earth Observation Metadata profile of Observations & Measurements," OGC, 2016.

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

[31 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/.

[32 GEO, "GEOSS," [Online]. Available: http://earthobservations.org/geoss.php.

[33 ESA, "About TEP," 2017. [Online]. Available: https://tep.eo.esa.int/about-tep.

[34 Copernicus, "The upcoming Copernicus Data and Information Access Services (DIAS)," 2017. [Online]. Available: http://copernicus.eu/news/upcoming-copernicus-data-and-information-access-services-dias.

[35 European Commission, "Functional Requirements for the Copernicus Distribution Services and the Data and Information Access Services (DIAS)," 2016.

[36 European Commission, "Annex to the Commission Implementing Decision on the Adoption of the Work Programme 2018 and on the Financing of the Copernicus Programme," 2018.

[37 GEOEssential Consortium, "Consortium Agreement for GEOEssential Project," 2017.

[38 ECOPOTENTIAL Consortium, "Data Management Plan," 2016.

[39 ECOPOTENTIAL Consortium, "Description of Work," 2015.

[40 A. Folino, A. Caruso and G. Aracri, "Semantic Services - GEOEssential Deliverable D1.4 release 2," 2019.

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

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

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

[44 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].

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

[46 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.

[47 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.

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

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

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

8 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.

9 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).

9.1 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.

9.2 Accessibility

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

9.3 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.

9.4 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.

9.5 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.

10 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.