ENERGIC OPEN DATA

European NEtwork for Redistributing Geospatial Information to user Communities - Open Data

This project is funded by the European Union

Project Details

Project reference:
620400

Funded under:
CIP

Duration:
36 months
01/10/2014 -
30/09/2017

www.energic-od.eu

ENERGIC OD

European NEtwork for Redistributing Geospatial Information to user Communities - Open Data

D5.1: VIRTUAL HUBS - SYSTEM ARCHITECTURE

This project is partially funded under the ICT Policy Support Programme (ICT PSP) as part of the Competitiveness and Innovation Framework Programme by the European Community

Document Information

Project title	European NEtwork for Redistributing Geospatial Information to user Communities - Open Data
Call identifier	CIP-ICT-PSP-2013-7
Project acronym	ENERGIC OD
Starting date	01.10.2014
End date	30.09.2017
Funding scheme	Competitiveness and Innovation Framework Programme (CIP) - ICT Policy Support Programme - Pilot B
Contract no.	620400

Deliverable no.	D5.1
Deliverable name	Virtual hubs - system architecture
Work Package	5
Nature[1]	R
Dissemination[2]	PU
Editor	Paolo Mazzetti (CNR-IIA)
Authors	Paolo Mazzetti (CNR-IIA) Stefano Nativi (CNR-IIA)
Contributors	Michael Bauer (GeoKomm) Stefan Braumann (LUP) Raffaella Brumana (POLIMI) Mattia Previtali (POLIMI) Miguel Àngel Latre Abadia (UNIZAR) Michael Müller (AED-SICAD)
Date	31-03-2015

[1] R = Report, P = Prototype, D = Demonstrator, O = Other

[2] PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services).

ENERGIC OD Consortium

National Research Council of Italy (CNR-IIA)
Italy

AED-SICAD Aktiengesellschaft (AED-SICAD)
Germany

Association Francaise pour l'Information Geographique (AFIGEO)
France

Bureau De Recherches Geologiques et Minieres (BRGM)
France

Global Infotech (NOW)
Italy

Centre National de la Recherche Scientifique (CNRS)
France

Depth France Sas (DEPTH)
France

Verband der GeoInformationswirtschaft Berlin/Brandenburg (GEOkomm)
Germany

Instytut Geodezji i Kartografii (IGIK)
Poland

Luftbild Umwelt Planung GmbH (LUP)
Germany

Politecnico di Milano (POLIMI)
Italy

Gesellschaft für Stadt-und Regionalplanung mbH (SRP)
Germany

Regione del Veneto (VEN)
Italy

Universidad de Zaragoza (UNIZAR)
Spain

Alkante Sas – Geonet (ALKANTE)
France

Revision History

Version	Date	Modified by	Comments
0.1	12.01.2015	Paolo Mazzetti (CNR-IIA), Stefano Nativi (CNR-IIA)	Preliminary Table of Content
0.2	13.01.2015	Paolo Mazzetti (CNR-IIA)	Section 2
0.3	26.01.2015	Paolo Mazzetti (CNR-IIA)	Section 3
0.4	13.02.2015	Paolo Mazzetti (CNR-IIA), Raffaella Brumana (POLIMI)	Updates from the WP4/WP5 workshop
0.5	20.02.2015	Paolo Mazzetti (CNR-IIA)	Section, 4 and 5. Complete revision
0.6	05.03.2015	Paolo Mazzetti (CNR-IIA), Michael Bauer (GeoKomm)	Update of section 2.3 with grades of openness
0.7	10.03.2015	Paolo Mazzetti (CNR-IIA)	First complete draft
0.8	16.03.2015	Paolo Mazzetti (CNR-IIA), Stefan Braumann (LUP) Miguel Àngel Latre Abadia (UNIZAR)	Updates of section 3 and 5 with content from D2.3 and D6.1 drafts
0.9	17.03.2015	Paolo Mazzetti (CNR-IIA), Stefano Nativi (CNR-IIA)	General revision
1.0	19.03.2015	Paolo Mazzetti (CNR-IIA)	Draft to be circulated to partners
1.1	28.03.2015	Mattia Previtali (POLIMI) Miguel Àngel Latre Abadia (UNIZAR) Michael Müller (AED-SICAD)	Comments to v1.0
2.0	07.04.2015	Paolo Mazzetti (CNR-IIA)	Second revision and proofreading

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS AND ACRONYMS

Abbreviation / Acronym	Description
CA	Consortium Agreement
EC	European Commission
EC-GA	European Commission Grant Agreement
TL	Task Leader
WP	Work Package
WPL	Work Package Leader

EXECUTIVE SUMMARY

The ENERGIC OD project aims *to build Virtual Hubs to facilitate the use of open geospatial data*. Unfortunately, most of the concepts in this apparently clear statement of objectives are controversial. There are many different interpretations of geospatial data and open data, and there is not any clear definition of what a Virtual Hub is. To guide the design of the ENERGIC OD architectural framework, we considered that the real keyword in ENERGIC OD is "facilitate", and everything should then be interpreted in that direction. For this reason, for example, ENERGIC OD cannot adopt a strict definition of open data limiting to "free-of-charge" data. ENERGIC OD has to facilitate users lowering barriers to open data such as the existence of heterogeneous interfaces, variety of formats, including proprietary ones or different coordinate reference systems.

The ENERGIC OD Virtual Hub concept puts its basis on past experiences in building System of Systems through a brokering approach. In brokered architectures, dedicated components provide mediation and harmonization of interfaces and data models avoiding the need of changes in the data provider systems.

As an innovation action, ENERGIC OD focuses on loosely-coupled integration of mature technologies and tools, most of them provided or under control of ENERGIC OD Consortium members. In particular, existing brokers – such as the GI-suite Brokering Framework adopted in the Global Earth Observation System of Systems – assure the basis to build advanced Virtual Hubs.

The integration of tools in the Virtual Hub will be based on full server-side APIs, while applications development will be facilitated through simple client-side APIs based on widespread Web technologies (HTML5, Javascript and CSS).

For greater flexibility, ENERGIC OD will adopt an agile methodology allowing rapid development in response of new requirements. It will have three main yearly iterations with fixed objectives for demonstration in reviews and events.

ENERGIC OD will deploy five national VHs in France, Germany, Italy, Poland and Spain. However, the architecture is flexible and can accommodate different topologies, to address specific requirements, such as the need of a sixth European VH, or a central VH acting as a single-point-of-access, as it could be suggested by marketing reasons for better exploitation. By a technical point-of-view, the deployment will be possible on local infrastructures, possibly managed by one or more ENERGIC OD partners, or on commercial or public clouds providing Infrastructure-as-a-Service (IaaS) and Platform-as-a-Service (PaaS) functionalities.

1 INTRODUCTION

This document describes the system architecture of the ENERGIC OD Virtual Hubs for facilitating access to Open Data. The document is released at project month 6, and therefore it was prepared when some specific aspects such as a survey of Open Data infrastructures (from WP3), the analysis of the state-of-the-art (from WP2), user requirements (from WP4), and specific requirements from applications (WP6), according to the DoW, were still under investigation by the Consortium. Therefore it cannot and does not aim to describe a low-level software architecture. Instead, it aims to provide a high-level view of the general approach, and the architectural framework in ENERGIC OD. It does not target software developers, but open data community members, interoperability experts, including non-technical members of the ENERGIC OD Consortium.

A first section focuses on the objectives and rationale behind the project, attempting to clarify the main relevant concepts in ENERGIC OD, such as what Open Data are, and providing a definition of Virtual Hubs.

A second section reports a preliminary analysis on actors, user requirements and system requirements. For the reasons reported above, it cannot be exhaustive, but it is sufficient to elicit the architectural principles.

The third section actually describes the ENERGIC OD architectural principles, focusing specifically on the need of loosely coupled applications, and on the brokering approach which is at the core of the ENERGIC OD Virtual Hub concept.

The fourth section provides an overview of the ENERGIC OD software system architecture according to the viewpoint modelling approach using the five views defined by the Reference Model for Object Distributed Processing from ISO.

A fifth section introduces the agile development approach that will be adopted by the ENERGIC OD project, and the sixth and final section reports short information on the deployment plan.

2 RATIONALE AND MAIN CONCEPTS

2.1 The ENERGIC OD main objective

The ENERGIC OD project is funded in the Competitiveness and Innovation Framework Programme and ICT Policy Support Programme (CIP ICT PSP). In particular it was proposed as a response to the call CIP-ICT-PSP-2013-7 for Pilots in the Objective 2.2: Open Data - Obj 2.2.a: "Open Data experimentation and innovation building on geographic information".

The key statement in the call says that:

> *The pilots should focus on the development of virtual hubs that facilitate the use of open (freely available) geographic data from different sources for the creation of innovative applications and services*

It identifies the goal of the project as *facilitate the use of open (freely available) geographic data* and the operational objective as the *development of virtual hubs*. The present document describes the system architecture for reaching the operational objective.

2.2 Geoinformation

As the CIP ICT PSP call requires, ENERGIC OD focuses on Geographic Information (GI) as "*an important means for creating innovative services*". Geographic Information is "*information concerning phenomena implicitly or explicitly associated with a location relative to the Earth*" [1]. 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*" [2].

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

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

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

- *Data (and information) producers* are mostly focused on data and metadata models and formats. Multiple standards have been defined addressing issues specific for different disciplinary domains such as HDF, netCDF and GRIB for EO data, ESRI Shapefile or OGC GML for feature type information. Proprietary formats are still widespread;
- *Data providers* are mainly focused on data sharing services. As for data models and formats, several standards have been designed and adopted in different disciplinary domains. For example, in the biodiversity context TDWG standards are widely adopted, in the meteo-ocean community THREDDS Data Server is a widespread technology. OGC standard services are commonly adopted in the GIS community. Light specifications like KML (now an OGC standard) or OpenSearch are also common. OAI-PMH is a standard for long-term preservation archives.
- *Overarching initiatives* influence technological aspects in several ways, in particular on data management (e.g. 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 CIP ICT PSP call explicitly mentions it saying that "*one of the main obstacles is the lack of agreed interoperability standards*". Unfortunately, as it will be explained later, the lack of interoperability standards is indeed an issue, but it is actually more the consequence of the complexity of the geospatial world than the reason of it. Including many actor categories, many disciplines, and many stakeholders (public authorities, private companies, citizens, etc.) **the complexity of the geospatial world makes impossible to agree on a single (or a small set) of standards** and, later, impose and enforce their adoption.

2.3 Open Data in ENERGIC OD

It is recognized that there is a lack of clarity about key terms in literature and public debates related to Open Data [3]. 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 [4].

Concerning ENERGIC OD, the call provides few hints limiting the scope to "*open (freely available) geographic data*" [5]. Although this definition helps to clarify the data typology (i.e. geographic data), it actually reiterates the gratis vs. libre ambiguity concerning policy: it does not specify whether "free" should be meant as "with little or no restriction" (*libre*) or "for zero price" (*gratis*).

The ENERGIC OD context is made even more complicated when the call refers to some examples of data: "*The aim is to stimulate innovation and business activities around GI data (including large dynamic European datasets <u>such as GMES data</u>)*". Indeed, for example, for GMES (now Copernicus) space component, it is stated that "*in addition to the data produced by the Sentinels satellites, Copernicus users can also have access <u>under certain conditions</u> to the data produced by other satellite missions* *referred to as 'Contributing Missions'"* [6]. The existence of conditions to access, in general, may have a strong impact on ENERGIC OD since it aims to "*Facilitate market entry of new companies, and the development of innovative services*".

As part of the WP3 activities, ENERGIC OD has decided to adopt an operational definition for Open Data [7] which bases on the Open Definition stressing that data is freely accessible but not exclusively free of charge.

Taking into account possibly complex use-cases like Copernicus one, the ENERGIC OD establishes a light definition of Open Data:

ENERGIC OD Open Data are geospatial data which can be accessed, used, modified, and shared for any purpose without unnecessary technological, financial and legal impediments.

This means that data can be considered "open" for the ENERGIC OD purposes if there is not any arbitrary or undocumented condition to access them. Obviously there are different grades of how open data actually is. To represent this, ENERGIC OD will make use of a rating system for evaluating the openness of a platform. Tim Berners-Lee suggested a 5-star deployment scheme for Open Data, which – with minor adaptations – can be applied to the platforms included in the ENERGIC OD Open Data inventory (Figure 2).

	make your stuff available on the Web (whatever format) **NOT** under an open license
★	make your stuff available on the Web (whatever format) **under** an open license
★★	make it available as structured data (e.g. vector data instead of a scan of a map)
★★★	structured non-proprietary formats (e.g. gml instead of shapefile)
★★★★	use URIs to denote things, so that people can point at your stuff
★★★★★	link your data to other data to provide context

Figure 2 - Tim Berners-Lee's 5-star Open data deployment scheme for Open Data adopted in ENERGIC OD

2.4 Virtual Hub

The call does not specify what a Virtual Hub is. The term is just used in a generic way saying that "*The pilots should focus on the development of virtual hubs that facilitate the use of open (freely available) geographic data*". The term *virtual hub* is usually adopted referring to the *hub-and-spoke* distribution paradigm, where it means "*Of or being a system of distribution, as of goods, passengers, or data, in which the items being distributed are routed into and out of a central location*" [8]. A software architecture adopting the *(message) broker* pattern is usually referred as a *hub-and-spoke* architecture [9].

The ENERGIC OD Description of Work (DoW) document describes the ENERGIC OD approach to Virtual Hubs as follows:

> *ENERGIC OD will deploy a set of Virtual Hubs (VH) by integrating an existing broker framework with other selected technologies to provide users with a single point of access to geospatial datasets provided by new or existing platforms and infrastructures, including INSPIRE-compliant systems and GMES/Copernicus services.*

The proposal built on those definitions and requirements, saying that a ENERGIC OD Virtual Hub "*Thanks to the brokering framework it is able to interconnect heterogeneous infrastructures and systems*" making the user "*able to seamlessly access geo-information from heterogeneous infrastructures*".

We provide then the following definition of an ENERGIC OD Virtual Hub:

An ENERGIC OD Virtual Hub is a virtual node where users can seamlessly access potentially unlimited datasets by brokering heterogeneous open geospatial data sources.

3 ANALYSIS

3.1 Actors

ENERGIC OD identifies a set of Actors, which is a set of user categories involved in: a) the setup and operation of Virtual Hubs, b) in the use of Virtual Hub resources, and finally, c) in the use of applications based on Virtual Hubs.

Actor	Acronym	Description
Virtual Hub Provider	VH Provider	The VH Provider is the person/organization that provides the VH capacities. Typically, it is a service provider that makes business in providing the VH capabilities to different users, including application developers.
Virtual Hub Administrator	VH Admin	The VH Admin is the person who manages a Virtual Hub configuring it for VH users and providing support.
Virtual Hub User	VH User	The VH User is a person who accesses the VH through the VH portal
Virtual Hub App Developer	VH App Developer	The VH App Developer is a person who develops and manages applications based on the VH APIs, and that access information through the VH
Virtual Hub Consumer	VH Consumer	A VH Consumer is a person who makes use of VH capabilities, which is either a VH User or a VH App Developer (or both).
Virtual Hub App End-User	End-User	An End-User is a person who makes use of an applications developed and/or provided through the VH. He/she is not necessarily aware of the existence of the VH.

3.2 User requirements

In the ENERGIC OD work plan, a specific Work Package (WP4) is dedicated to "Requirements and specifications: SDI, data harmonisation and applications addressing user needs".

ENERGIC OD User Requirements are collected from different sources:

a) Call text
b) ENERGIC OD DoW
c) Activities in WP6 “Development of new innovative applications”
d) Previous work in relevant initiatives and programmes at national, regional, European and international level (including Copernicus, INSPIRE, GEOSS)
e) Survey of stakeholders needs

In terms of user requirements, at this stage the ENERGIC OD Virtual Hub was conceived as a typical data sharing system with a specific focus on solving interoperability issues to facilitate usage of open data. Indeed, no specific functionality is needed beside those required to support the typical data sharing scenario shown in Figure 3, with the general functionalities of Discovery (supporting search for relevant resources), Evaluation (supporting inspection of resources to evaluate value and relevance), Access (supporting retrieval of relevant resources) and Use (from simple visualization to complex processing where required). However 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 3 The typical high-level scenario in the geospatial domain

The deliverable D6.1 “Application based requirements and standards catalogue” detailed the specific requirements for the ten pilot applications developed in ENERGIC OD. It provided indication of a fifth high-level user requirement which is Publishing, since some of the applications need to store generated or acquired datasets.

The release of the D6.1 “Application based requirements and standards catalogue” will allow to refine user requirements in the next iterations (see §6)

3.3 System Requirements

In the ENERGIC OD work plan, a specific Work Package (WP4) is dedicated to “Requirements and specifications: SDI, data harmonisation and applications addressing user needs”.

ENERGIC OD System Requirements are collected from different sources:

a) Call text
b) ENERGIC OD DoW
c) Elicitation from user requirements (Section §3)

Actually the elicitation from user requirements will be possible only after the release of the ENERGIC OD deliverable D4.2 “Report on the most relevant fields of application and user needs, strictly related to innovative applications, services and semantic content to be implemented” planned at project month 12. However preliminary discussions have been already carried out. In particular, the joint WP4 and WP5 workshop held on February 2015 in Florence provided several hints.

Table 1 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;
FR2	Semantic discovery	The system provides semantic enhancements for discovery, supporting multilingualism, suggestions, and search for related terms.
FR3	Dataset access	The system provides access to datasets from heterogeneous data provision systems
FR4	Dataset transformation	The system supports basic transformation functionalities such as: a) subsetting b) interpolation c) reprojection on multiple Coordinate Reference Systems d) data format 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.). Relevant transformations will be identified during the project (a survey dedicated to this aspect is planned in WP4 activities).
FR5	Support of legacy/proprietary formats	The system must be able to support a set of relevant legacy/proprietary formats. This is necessary to support relevant open data source with low rate of openness (see §2.3). Relevant formats will be identified during the project (a survey dedicated to this aspect is planned in WP4 activities).
FR6	Data processing	The system must be able to process data. Only general processing, that is, processing required by multiple applications, will be supported by the system (a survey dedicated to this aspect is planned in WP4 activities). Specific processing, tailored to single applications is up to the application developer.
FR7	Data source monitoring	The system must provide feedback on the status of remote data sources.
FR8	AAA	The system must support Authentication, Authorization

		and Accounting allowing collecting information about the use for both technical and marketing purposes.
FR9	Data Publishing	The system must support data publishing from data collection applications
NFR1	Seamless discovery and access	The system provides discovery (FR1) and access (FR2) of heterogeneous data sources through any of multiple standard interfaces
NFR2	Multiple distribution strategies	The system can be replicated on multiple instances depending on technological or marketing choices.
NFR3	Multiple deployment strategies	The system can be deployed according to different strategies. In particular it can be deployed either locally (on local infrastructures) or on commercial/public Infrastructure-as-a-Service (IaaS) clouds.
NFR4	APIs	The system functionalities must be accessible both server-side (for integration of tools enhancing system capabilities) and client-side (for application development through mash-up)
NFR5	Availability	The system must assure high availability
NFR6	Performance	The system must assure adequate performances
NFR7	Scalability	The system must assure adequate scalability in terms of number of data sources, number of users, number of requests, etc.
NFR8	Security	The system must assure security
NFR9	Usability	The system must be user-friendly for both end-users and application developers
NFR10	Extensibility	The system must be extensible to support new data sources protocols, new apps without major changes

Table 1 ENERGIC OD system requirements

4 ENERGIC OD ARCHITECTURAL PRINCIPLES

4.1 Open Software Architectures

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, a Virtual Hub must be conceived 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 VH is a particular intermediate component that facilitates the connection between end-user applications and data sources, contributing to the ecosystem evolution itself.

Living in an ever-changing context, the VH must be also able to evolve in response to those changes. Indeed, although the VH requirements can be clear at this stage of the ENERGIC OD project, in order to support the sustainability of outcomes, it is necessary to assure that the VH architecture and implementation can (easily) evolve.

Software evolution has been the subject of several research works in the past (Table 2). A first classification [10] 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 2 Different categories of techniques to support software evolution

It is quite evident that a centralized evolution model is not an option for the ENERGIC OD VH for several reasons: a) a VH is not fully based on software which is under control of a single organization; b) even the ENERGIC OD 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 ENERGIC OD Consortium could achieve the role of central authority, it exists only until the end of the project, while the sustainability of VH must be considered also beyond the ENERGIC OD project lifetime.

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

For the ENERGIC OD 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 (see also the Consortium Agreement establishing that the background software is provided at the minimum level of interaction required for the use in the project – ranging from running instances as Software-as-a-Service, to the executable code, and finally source code); 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 ICT PSP supports only *technical adaptation and integration work*, encouraging 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 **ENERGIC OD Virtual Hubs 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.

4.2 Brokered Systems of Systems

4.2.1 System of Systems Engineering

Taking into account the main constraint of the ICT PSP: "*The ICT PSP <u>does not support research activities</u>; it may cover, when needed, technical adaptation and integration work in order to achieve the objectives*" (underline in the original text), ENERGIC OD bases on the Open Architecture paradigm.

Interoperability is recognized as the main challenge for ENERGIC OD. As the call says: "*Solutions should lead to an <u>easier discoverability and use</u> of geographic information available for use in innovative applications and services, and where possible <u>draw together datasets from different sources</u>*". The ENERGIC OD proposal is based on the successful experience of brokered architectures to implement Systems of Systems.

The notion of "System of Systems" (SoS) and "System of Systems Engineering" (SoSE) emerged in many fields of applications to address the common problem of integrating many independent, autonomous systems, frequently of large dimensions, in order to satisfy a global goal while keeping them autonomous. Therefore SoSs can be usefully described as follows: *systems of systems are large-scale integrated systems that are heterogeneous and consist of sub-systems that are independently operable on their own, but are networked together for a common goal* [11]. It is evident that this definition fits well in the ENERGIC OD context where sub-systems like the INSPIRE infrastructure, Copernicus core and downstream services are clearly out of control of the ENERGIC OD Consortium, and even from possible future exploitation scenarios.

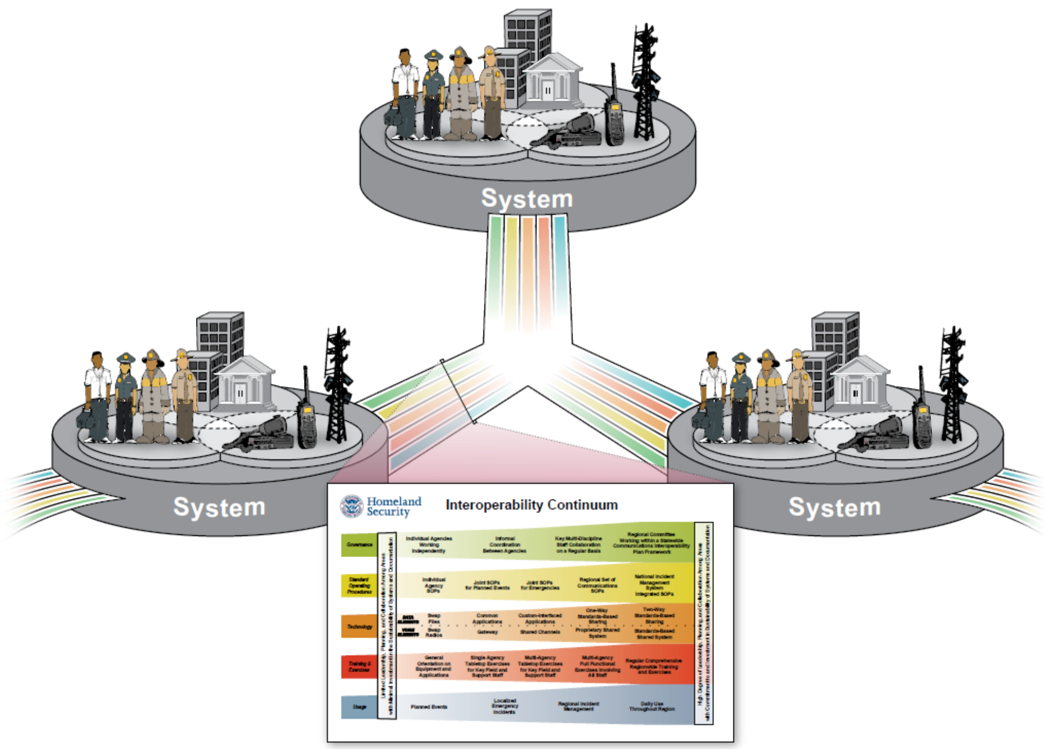

Figure 4 System of Systems in Practice – from [12]

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* [13], 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" [14]. 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" [14]. 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 happens 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, 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 in web developers, and often in 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 while 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 of interoperability – 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 presents 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 noteworthy, that the former is currently addressed by the Brokering Interest Group[3] 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+brokered approach.

Standardization helps to reduce the redundant heterogeneity, while brokering addresses the remaining irreducible heterogeneity.

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

a) There are several data sources of interest for ENERGIC OD 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)[4]

b. In the meteo-ocean community, the UNIDATA THREDDS Data Server (TDS)[5] is widely adopted

c. Many Open Data communities share the CKAN[6] technology for implementing data portals.

b) ENERGIC OD 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 in the problem of allowing M different applications to interact with N different data sources. This is an MxN complexity problem. 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 simply, 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 interacting with each server) to M+N (each client and each server only need to interact with the brokers).

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

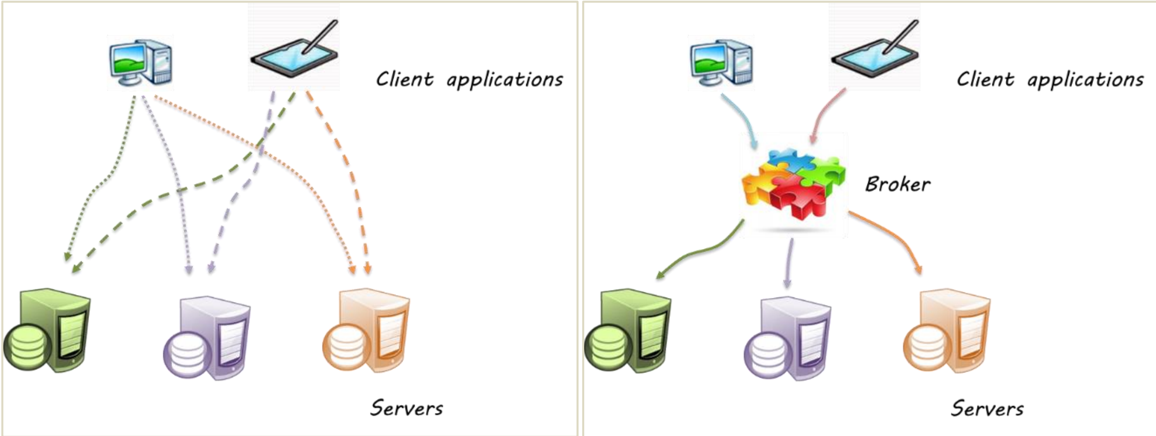

Figure 5 Client-Server vs. Brokered Architectures

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. ENERGIC OD 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. In order to allow provision of resources from Virtual Hub providers who do not contribute to any sub-system, a specific (local) sub-system will be set up.

4.3 ENERGIC OD service provision model

In the recent 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*" [15].

The cloud model includes three different kinds of services [15]:

- *Infrastructure as a Service (IaaS)*: The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. Examples are Amazon Elastic Compute Cloud (EC2) and Amazon Simple Storage Service (S3).
- *Platform as a Service (PaaS)*: The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages, libraries, services, and tools supported by the provider.
- *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 aiming to support new markets and new businesses as in ENERGIC OD. Indeed, it presents some advantages: a) it widen the range of users, requiring only a browser and a good connectivity which is currently easy to achieve even in mobility, b) it separates responsibilities, delegating support services (hardware and software management, accounting and billing) to cloud providers, and allowing developers to focus on their own application.

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

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

The VH 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

ENERGIC OD requirements can be broadly classified into two categories:

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

The general ENERGIC OD 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 ENERGIC OD 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 ENERGIC OD Architectural principles

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

P1. ENERGIC OD Virtual Hubs adopt an Open Software Architecture

P2. ENERGIC OD Virtual Hubs are developed integrating and adapting existing software solutions

P3. ENERGIC OD Virtual Hubs adopt a Decentralized Software Evolution

P4. ENERGIC OD Virtual Hubs are made of software components interacting through (low-level) APIs

P5. ENERGIC OD Virtual Hubs are the common infrastructure of a brokered System of Systems

P6. ENERGIC OD Virtual Hubs expose a set of (high-level) APIs for interaction with the external environment

P7. ENERGIC OD Virtual Hubs are accessible according to the Software-as-a-Service (SaaS) and Platform-as-a-Service (PaaS) models, for end-users and developers respectively

P8. ENERGIC OD Virtual Hubs can be deployed either on private infrastructures or commercial or public clouds providing Infrastructure-as-a-Service (IaaS) capabilities.

P9. ENERGIC OD security architecture is orthogonal to the ENERGIC OD resource-sharing architecture.

5 ENERGIC OD 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*" [16]. 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*" [17].

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*" [17].

The following paragraphs provide the ENERGIC OD Virtual Hub description according to the following main views adopted in the ISO Reference Model for Open Distributed Processing [18]:

- 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;
[18]

The enterprise viewpoint focuses on the actors, and their interactions in scenarios and use-cases. The ENERGIC OD main actors have been described in §3.1. A detailed analysis of application use-cases and scenarios is planned in WP6, and it is documented in D6.1 "Application based requirements and standards catalogue".

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.
[18]

Figure 6 shows the ENERGIC OD layered architecture. It includes the following layers:

- **Data Access** layer: this layer provides data discovery and access functionalities to heterogeneous data systems.

- **Data Harmonization** layer: this layer provides harmonized discovery and access to heterogeneous data systems. Above this layer, the heterogeneity of data sources is hidden: they appear as a single data source.
- **Data Processing** layer: this layer enriches discovery and access with processing and semantics services.
- **User Interface** layer: this layer provides user-friendly access to data for both end-user (Graphical User Interface) and application developers (software libraries).

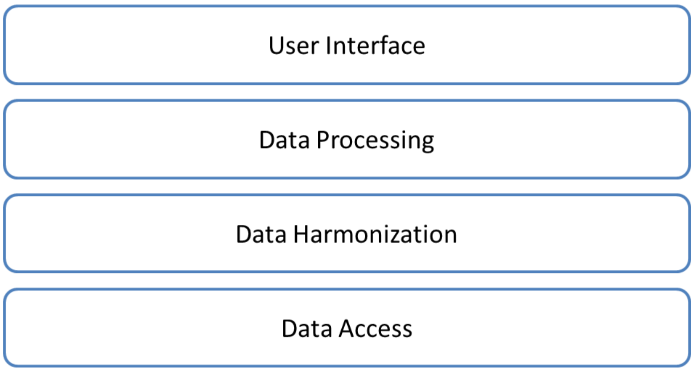

Figure 6 ENERGIC OD layered architecture

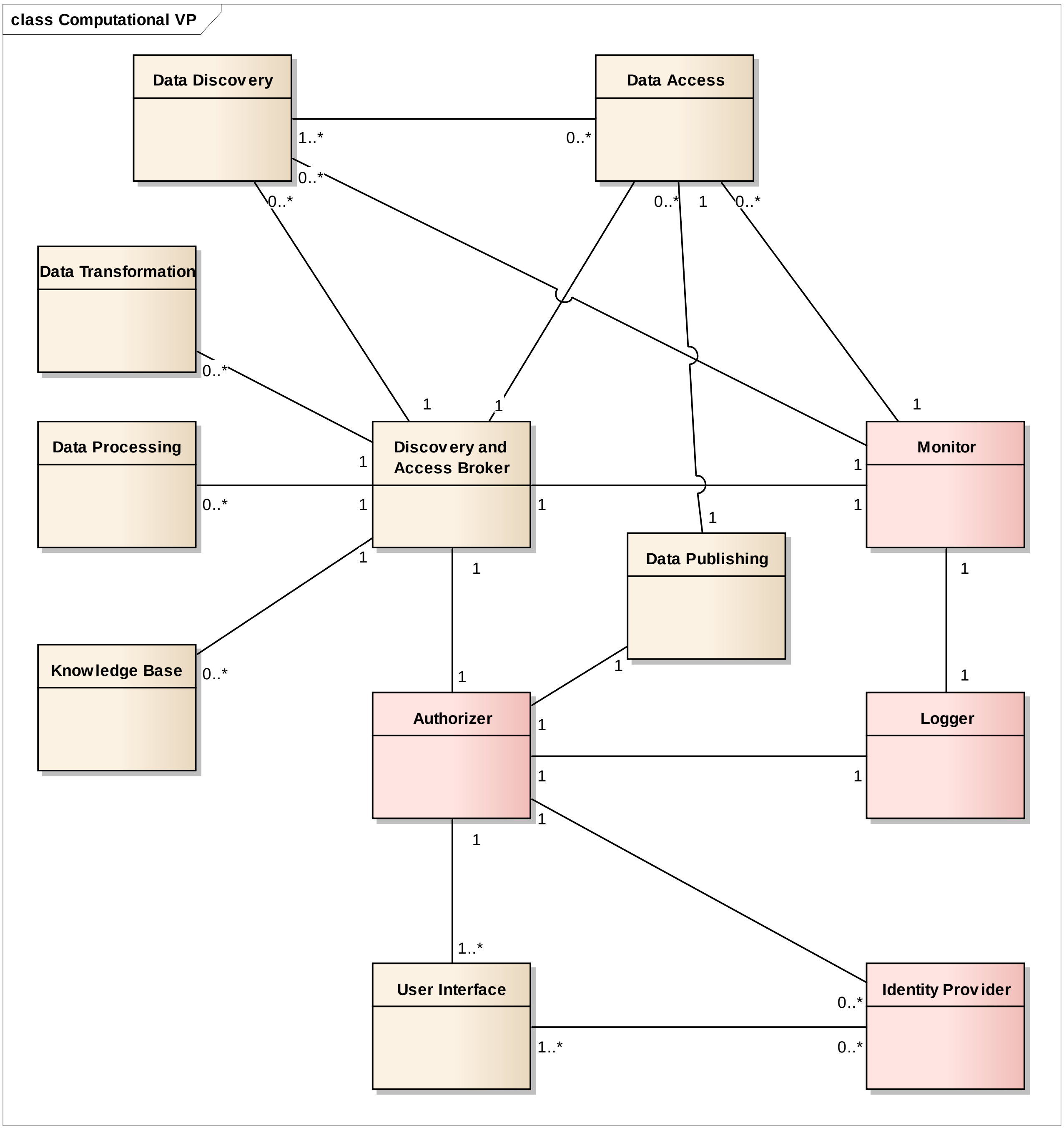

Figure 7 Main components of the ENERGIC OD architecture

Figure 7 shows the UML class diagram of the main functional components in the ENERGIC OD architecture. Components in red are not involved in the resource-sharing functionalities of the ENERGIC OD system (i.e. facilitate access to open geospatial data); they are either components of the security architecture or ancillary components improving the ENERGIC OD overall system capabilities.

The main functional components are described in Table 3 along with a reference to the functional and non-functional requirements they contribute to fulfill.

Component	Description	Relevant requirements
Data Discovery	It provides discovery of open data from specific data sources	FR1,
Data Access	It provides access to open data from specific data sources	FR3
Discovery and Access Broker	It accesses Data Discovery and Data Access components providing harmonized access to heterogeneous data sources. It must provide at least the following output interfaces and formats: • WMS, WFS, WPS • ASCII text file, PDF It must support queries by temporal extent, spatial coverage.	FR1, FR2, FR3, FR4, FR5, FR6, FR7 NFR1
Data Transformation	It transforms data changing resolution, Coordinate Reference System, format, etc. The content and semantic level of data is not changed.	FR4
Data Processing	It processes data to generate products through elaboration, fusion, and integration of datasets	FR6
Knowledge Base	It provides encoding of knowledge, to support advanced discovery services	FR2
Data Publisher	It provides publishing of data generated or collected by user applications. It must support the upload at least of: georeferenced text, images, data in general.	FR9
Authorizer	It checks if the user is authorized to perform an operation based on his/her identity and permissions	FR8 NFR8
Identity Provider	It checks the user's identity	FR8 NFR8
Monitor	It checks availability and status of data sources	FR7
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.	FR7,FR8 NFR8
User Interface	It handles the interaction between the user and the system. It includes GUIs allowing presentation of maps with pan and zoom, layer selection. It must support 2D maps and 3D landscape scenes. It must provide data and metadata s tables and charts.	NFR9

Table 3 ENERGIC OD main components

From Table 3 it is evident the core role of the Discovery and Access Broker which impacts on many functional and non-functional requirements. Moreover it appears that some requirements are not addressed (in particular non-functional requirements). Actually they are addressed mostly in the deployment architecture, therefore they will be discussed in the Engineering Viewpoint.

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.
[18]

As a project finalized to the creation of Virtual Hub facilitating the use of geospatial open data, the characteristics of information handled and shared by the system is a fundamental aspect.

ENERGIC OD addresses two main challenges concerning information handled by the Virtual Hub:

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

5.4.1 Heterogeneity

The ENERGIC OD Virtual Hub aims to facilitate the use of geospatial open data. As such it must take care of all the mediation, harmonization and transformation actions needed to make geospatial open data easily discoverable, accessible, and usable. Open data comes in many different shapes (see section §2.3), and ENERGIC OD cannot assume any kind of standardization.

This means that a Virtual Hub must be able to handle different service interfaces and metadata/data models for discovery and access. WP3 on "Open Data Survey" and WP4 on "Requirements and specifications: SDI, data harmonisation and applications addressing user needs" will provide the complete set of service interfaces and metadata/data models for discovery and access that the Virtual Hub must be able to connect as data sources and publish towards applications. Table 4 shows a preliminary list of required Virtual Hub protocols based on previous works and partially already supported by the ENERGIC OD core technologies (see section §5.6). They will be extended based on requirements from WP4 and WP6.

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	Discovery interface and metadata profiles

1.0, ebRIM/CIM, ebRIM/EO, CWIC	
FLICKR	Discovery and access interfaces
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 4 Preliminary list of Virtual Hub data sources protocols

Table 5 shows a preliminary list of required Virtual Hub protocols based on previous works and partially already supported by the ENERGIC OD core technologies (see section §5.6). They will be extended based on requirements from WP4 and WP6.

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 5 Preliminary list of protocols supported by Virtual Hub publishing interfaces

5.4.2 Semantics

The Virtual Hub addresses semantics through a query expansion strategy. When a query is submitted to the VH, the VH can ask external semantics services, to resolve keywords, providing "related" terms back. The returned concepts are used as keywords of multiple geospatial queries [19]. Then, the results from geospatial queries include responses not only to the original keywords but also to semantically related terms. (See Knowledge Base component in Figure 7, and Figure 8 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 the 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.

The GI-suite brokering framework, one of the core technologies for the VH implementation, already supports interaction with knowledge bases publishing SPARQL (SPARQL Protocol and RDF Query Language) interfaces to SKOS knowledge bases. It is operating in GEOSS tested with a knowledge base published by the EC-JRC and including the following knowledge bases:

- The General Multilingual Environmental Thesaurus (GEMET): 28 of the 31 languages currently provided by the EIONET portal.
- The INSPIRE Feature Concept Dictionary and Glossary: 21 of the 23 EU official languages for INSPIRE Themes, monolingual the other terms.
- The ISO 19119 categorisation of spatial data services: 21 of the 23 EU official languages.
- The GEOSS Societal Benefit Areas: 5 languages.

WP4 and WP6 will provide the requirements in terms of knowledge bases and related interfaces, that the VHs shall support.

5.5 Engineering Viewpoint

Engineering VP is concerned with the infrastructure required to support system distribution. [18]

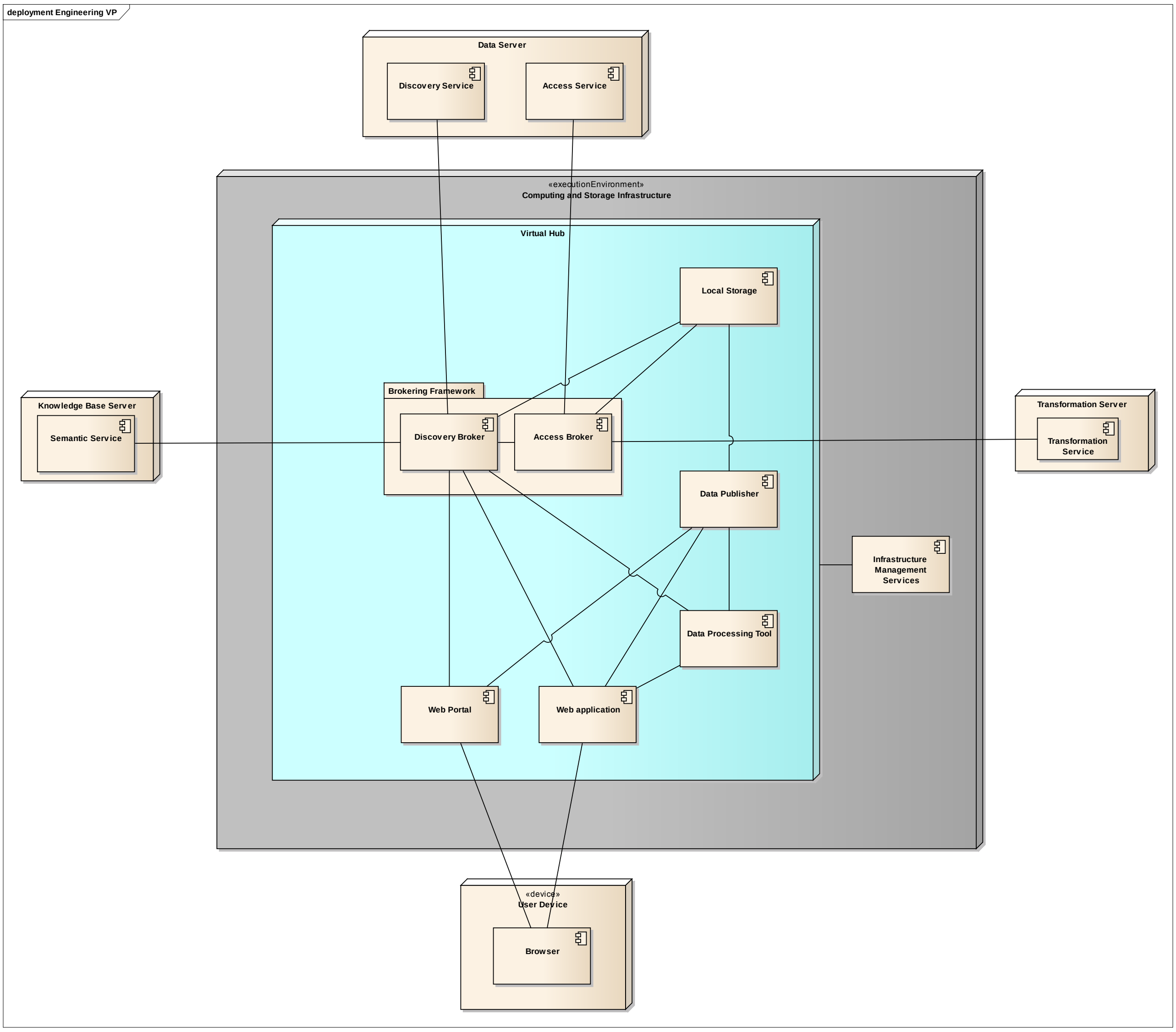

Figure 8 Engineering view of the ENERGIC OD architecture

Figure 8 shows the engineering view of the ENERGIC OD architecture. According to the architectural principle P9 the security and interoperability architecture can be decoupled, therefore, for the sake of clarity the security components are not shown there.

The ENERGIC OD architecture includes a set of different nodes:

Node	Description
Data Server	A Data Server is a node dedicated to serve open geospatial data. ENERGIC OD assumes that Data Server nodes are existing, up and accessible, and providing at least a discovery and an access service. No assumption is made about communication protocols.
Virtual Hub	A Virtual Hub is the core architectural node. It contains all the tools needed to achieve the ENERGIC OD objective of "facilitating the use of open geospatial data". In the ENERGIC OD proposed deployment there will be 5 Virtual Hubs at national level and one at European level.
Computing and Storage Infrastructure	A Computing and Storage Infrastructure is a node hosting one or more Virtual Hubs. It may be either a node managed locally by one ENERGIC OD partner, or a commercial or public cloud offering Infrastructure-as-a-Service capabilities.
Knowledge Base Server	A Knowledge Base Server is a node providing services accessible by the Virtual Hub for semantic enhancements. No assumption is made about communication protocols.
Transformation Server	A Transformation Server is a node providing services accessible by the Virtual Hub for data transformations (e.g. re-projection on different Coordinate Reference Systems, format encoding, sub-setting, change of resolution and interpolation) enhancements. 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 Web browser.

These nodes collectively host the software components interacting for an easier use of open geospatial data:

Component	Description
Brokering Framework	At the core of the Virtual Hub the Brokering Framework package includes a set of components which harmonize discovery and access of heterogeneous open geospatial data sources. It includes at least: • A Discovery Broker which connect with many different discovery, registry and inventory services, exposing several standard or well-known discovery interfaces. Through this well-known interfaces, a user can discovery all the datasets published by the different data sources. • Support for semantic enhancement of discovery. A simple query can be expanded in multiple queries based on the semantics relationship defined in an external knowledge base. • An Access Broker which connects with many different access

	and download services, exposing several standard or well-known access interfaces. Through these well-known interfaces, a user can access all the datasets published by the different data sources. • Support for data transformation. Multiple datasets can be transformed accessing external transformation services, in order to harmonize them on the same Common Grid Environment (same spatial and temporal coverage, same resolution, same Coordinate Reference System, same data format, etc.)
Semantic Service	Semantic Services expose knowledge-bases such as thesauri, gazetteers, ontologies, allowing to find terms related to a keyword for query expansion.
Transformation Service	Transformation Services implement datasets transformation (e.g. subsetting, re-projection, interpolation)
Local Storage	The Local Storage stores information generated by applications and by the Virtual Hub. It is seen as any other Data Source by the Brokering Framework
Data Publisher	The Data Publisher allows storing datasets in the Local Storage.
Data Processing Tool	A Data Processing Tool is a component that processes datasets retrieved through the Brokering Framework and store the results on the Local Storage.
Web Portal	A Web Portal is the primary interface for Human-to-Machine interaction. It allows at least discovery, upload and download of datasets for offline usage.
Web Application	A Web Application is a specific component implementing (part of) the application logic of a Web or mobile app. It implements the needed workflow interacting with the Brokering Framework, the Data Publisher, etc.
Browser	The Browser is the component enabling user's interaction with the system. It will host part of the application logic (as client-side code) and the presentation logic.

Figure 9 shows 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.

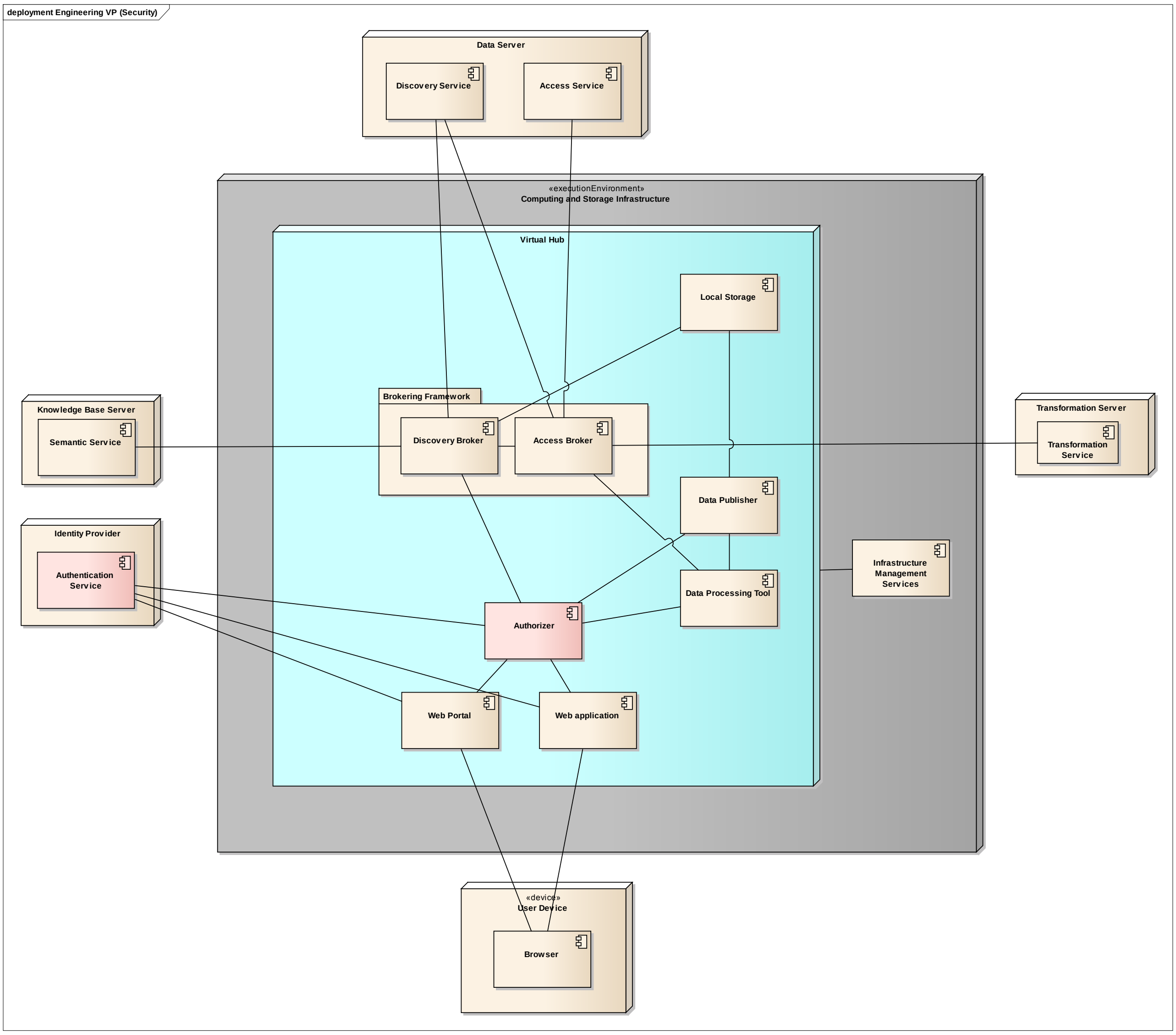

Figure 9 Engineering view of the ENERGIC OD architecture showing the main security components

5.6 Technology Viewpoint

Technology VP is concerned with the choice of technology to support system distribution. [18]

The ENERGIC OD system will be implemented using and extending existing solutions and tools. At the time of the proposal and DoW preparation some technologies were preliminarily identified. They were mostly provided or under control of ENERGIC OD partners. Table 6 shows the approach and available solutions for each of the main architectural components.

WP2 carries out an analysis of the state-of-the-art from past and on-going research and development projects. The deliverable D2.3 “R&D State-of-the-art report” includes reference to technologies potentially useful in ENERGIC OD. They will be considered in future iterations of the architecture.

Component	Approach	Available solutions
Brokering Framework	Existing+Integration+Enhancement	The GI-suite Brokering Frameworkimplements discovery and access brokering towards more than 40 different data source types, publishing more than 10 different interfaces. It supports both metadata harvesting and distributed queries, configurable per data source. It implements query expansion using external knowledge bases (SPARQL/SKOS interface and model). It implements dataset transformation using internal algorithms or external services (WPS interface). The GIS-Broker supports federation of SDIs for GIS data.
Semantic Service	Existing+Development if required	The semantic service published by EC-JRC and providing a set of aligned thesauri will be initially used for multilingualism, suggestions, and semantic queries. Other knowledge base can be developed and published using open source tools supporting SPARQL/SKOS
Transformation Service	Existing+Development if required	Subsetting will be supported through GI-suite internal implementation Simple interpolation schemes will be supported through GI-suite internal implementation Most used Coordinate Reference Systems will be supported. Specific CRS transformations will be implemented if needed. FUSION Data Service supports data and CRS transformations.
Local Storage	Integration	Several geospatial open source tools may be used to implement local storage
Data Publisher	Integration	SpatiumCube, GeoServer, Degree, PostgreSQL, EasySDI, and several geospatial open source tools may be used to implement data publishing. CatMDEdit enables metadata management Sync'Serv enables database synchronization
Data Processing Tool	Integration	OpenRoute enables path routing
Web Portal	Development	GIS-Portal, a Stack of Open Source

		software for geomatics, the Atl@nte platform, and several open source tools may be used to implement data portals
Web Application	Development (see WP6)	Web APIs already available for the GI-suite Brokering Framework GeoExt, OpenLayers and several other Javascript libraries are available to develop geospatial Web applications
Browser	Existing	All major browser will be supported using HTML5+CSS+widespread Web technologies
Authentication Service	Existing+Integration	Widespread solutions (OpenAuth, Facebook, Google+, LinkedIn) will be adopted
Authorizer	Integration+Development	Open source tools will be identified and adopted

Table 6 Technological approach and potential solutions

6 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 10) [20] 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) [21]:

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

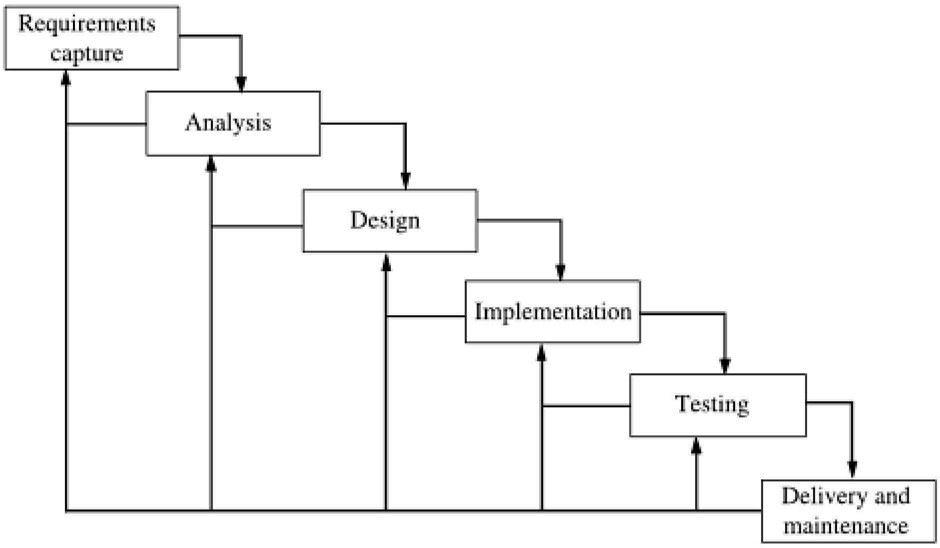

Figure 10 The traditional Waterfall Model (from [20])

Agile methodologies better respond to changes through an iterative process (Figure 11). 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.

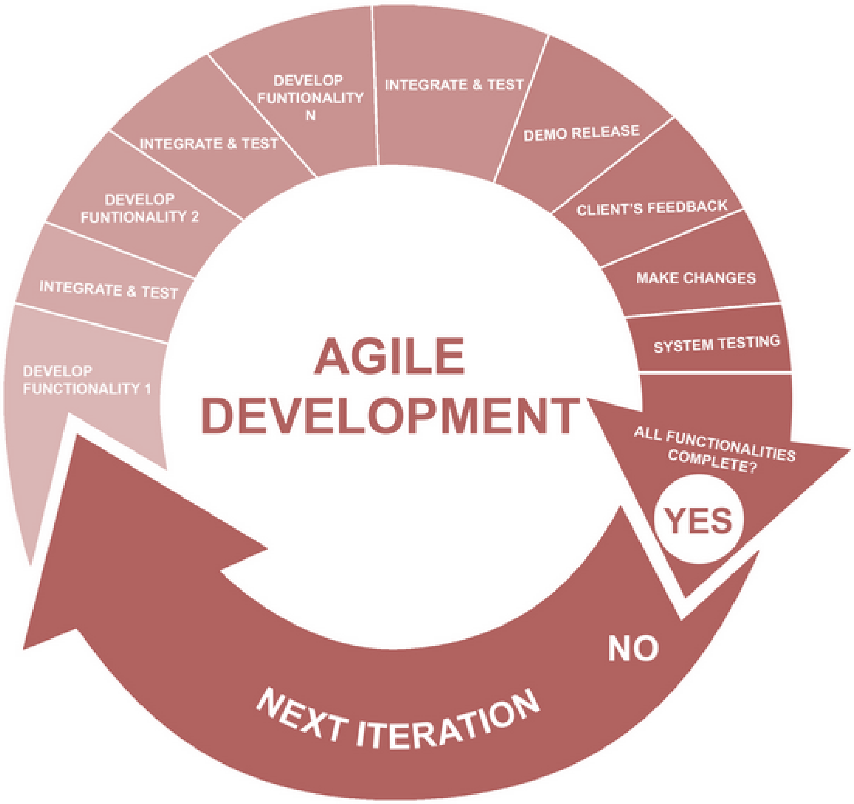

Figure 11 The iterative process in Agile development

Taking into account the specificity of ENERGIC OD we can identify three main milestones and therefore three main iterations:

- Project Month 12, first project review, end of the first major iteration and release of the Initial VH Capacity
- Project Month 24, second project review, end of the second major iteration and release of the Advanced VH Capacity
- Project Month 36, third and final project review, end of the third and last iteration and release of the Full VH Capacity

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 VH capacity

The first iteration will start at project-month 8 with T5.2 "Virtual Hub Implementation".

7 DEPLOYMENT CONSIDERATIONS

As stated in the ENERGIC OD DoW [22] the project will realize five national VH in France, Germany, Italy, Poland and Spain. Moreover, a sixth VH at European level can be optionally implemented if considered useful, possibly to aggregate the five national VH (Figure 12).

During an internal workshop held in Florence [23], the project partner leading WP8 on exploitation noted that, by a marketing point-of-view, it would be preferable to have a single point of access, instead of many. It would actually "facilitate" the user as required by the CIP ICT PSP call, implementing a one-stop-shop approach.

This possible requirement has actually a minor impact on the architecture, just stressing the importance to keep the architecture as loose as required to support many different topologies (see architectural principles in section §4.5).

The initial deployment will follow the DoW prescription, but different topologies will be later considered based on WP8 (and other WPs) inputs.

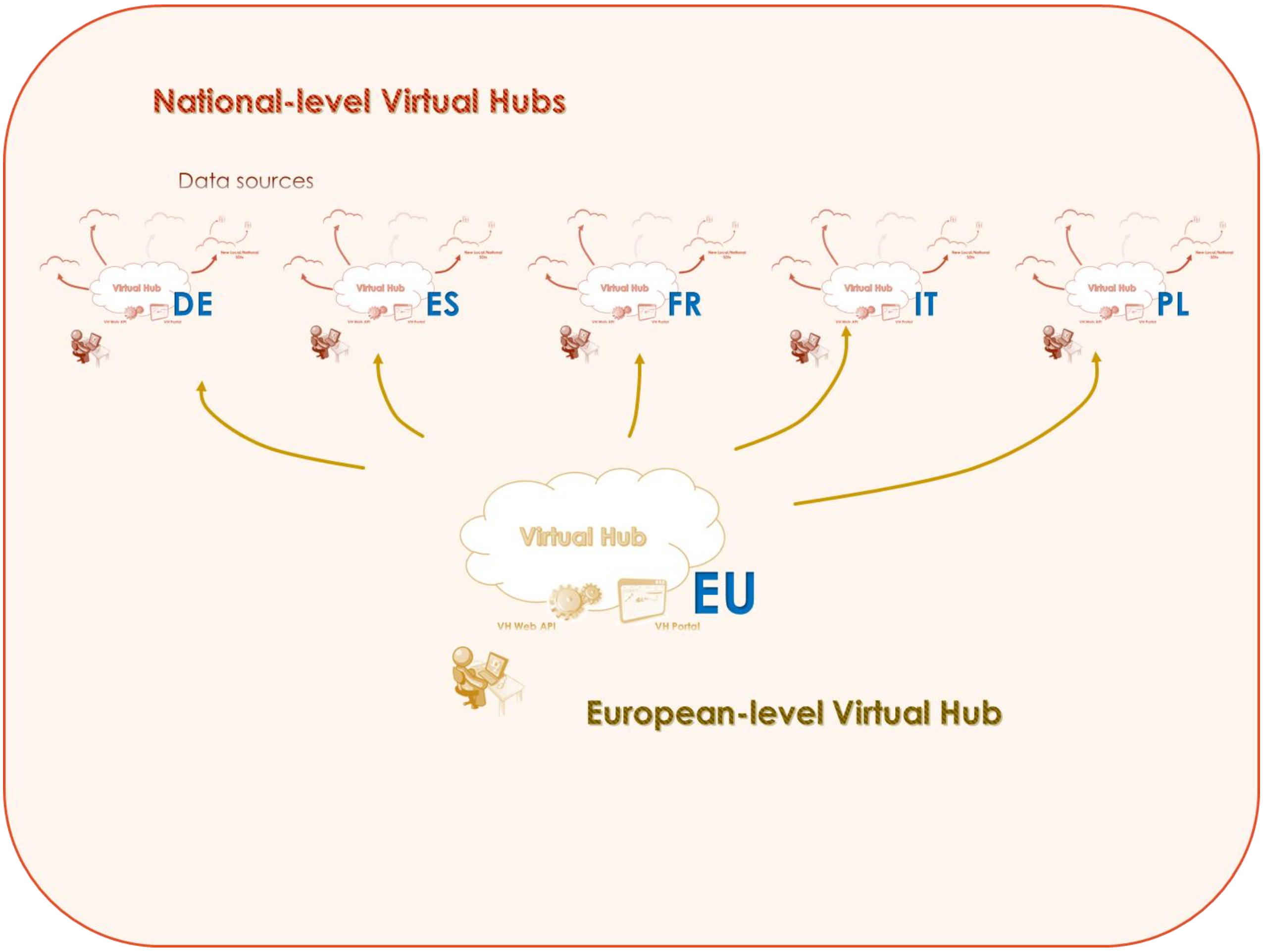

Figure 12 Deployment plan

8 REFERENCES

[1] ISO, «Geographic information — Reference model — Part 1: Fundamentals,» 2014.

[2] 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.

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

[4] «Open Definition,» [Online]. Available: http://opendefinition.org/od/.

[5] «Competitiveness and Innovation Framework Programme (CIP) ICT Policy Support Programme ICT PSP Work Programme,» 2013.

[6] Copernicus, «Copernicus Data Access,» [Online]. Available: http://www.copernicus.eu/main/data-access.

[7] J. Ernst, «Methodology and a list of criteria for the assessment of open data platforms and initiatives across the EU,» ENERGIC OD Consortium, 2015.

[8] American Heritage® Dictionary of the English Language - Fifth Edition, Houghton Mifflin Harcourt Publishing Company, 2011.

[9] F. J. Chang, Business Process Management Systems: Strategy and Implementation, CRC Press, 2005.

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

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

[12] Department of Homeland Security, «The System of Systems Approach for Interoperable Communications».

[13] S. Nativi, M. Craglia e J. Pearlman, «Earth Science Infrastructures Interoperability: The Brokering Approach,» *IEEE JSTARS,* vol. 6, n. 3, pp. 1118-1129, 2013.

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

[15] P. Mell e T. Grance, «The NIST Definition of Cloud Computing,» 2011.

[16] ISO/IEEE, «Systems and software engineering - Architecture description,» 2011.

[17] N. Rozanski e E. Woods, «Software Systems Architecture: Working with Stakeholders Using Viewpoints and Perspectives,» Addison-Wesley, 2011.

[18] ISO/IEC, «Information technology — Open Distributed Processing - Reference model: Overview,» ISO, 1998.

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

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

[21] 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 e D. Thomas, «Manifesto for Agile Software Development,» [Online]. Available: http://agilemanifesto.org/.

[22] ENERGIC OD Consortium, «Annex I - Description of Work,» 2014.

[23] ENERGIC OD Consortium, «Report on WP5 Workshop - D5.9,» 2015.

[24] S. Nativi, D. B. Khalsa Siri Jodha, M. Craglia, J. Pearlman, P. Mazzetti e R. Rew, «The Brokering Approach for Earth Science Cyberinfrastructure,» 2011.