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 (SECOND RELEASE)

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 second release
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) Olivier Frézot (BRGM) Michael Bauer (GeoKomm) Nargess Kamali (GeoKomm) Stefan Brauman (LUP) Helga Kuechly (LUP) Raffaella Brumana (POLIMI) Mattia Previtali (POLIMI) Miguel Ángel Latre Abadia (UNIZAR) Michael Müller (AED-SICAD)

[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).

	Tino Desjardins (SRP) Frank Iden (SRP)
Contributors	WP5 partners of ENERGIC OD Consortium
Date	29-02-2016

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
Trilateral Research Ltd. (TRI) - United Kingdom	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 GmbH (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	08.01.2016	Paolo Mazzetti (CNR-IIA)	Draft revision to be circulated for partners' contribution
0.2	02.02.2016	Paolo Mazzetti (CNR-IIA)	Added info on technological choices (GI-suite Brokering Framework)
0.3	03.02.2016	Mattia Previtali (POLIMI)	Integration of info on requirements from WP4
0.4	05.02.2016	Tino Desjardins (SRP), Frank Iden (SRP)	Info on integration of the GI-suite Brokering Framework and SRP GIS-Broker
0.5	08.02.2016	Miguel Angel Latre (UNIZAR)	Integration of info on technologies and research outcomes from WP2
0.6	09.02.2016	Paolo Mazzetti (CNR-IIA)	Harmonization of contributions
0.7	10.02.2016	Paolo Mazzetti (CNR-IIA)	Info on the GI-suite Brokering Framework
0.8	15.02.2016	Nargess Kamali (GeoKomm)	Revision of info from WP3. List of data sources for Virtual Hubs
0.9	15.02.2016	Paolo Mazzetti (CNR-IIA)	General revision
1.0	16.02.2016	Paolo Mazzetti (CNR-IIA)	Editorial changes. Second draft circulated to partners.
1.1	22.02.2016	Miguel Angel Latre (UNIZAR), Mattia Previtali (POLIMI)	Revision of contributions from WP2 and WP4
1.2	23.02.2016	Stefan Braumann (LUP), Helga Kuechly (LUP)	Integration of info from WP6 and specific info from the new version of D6.1
1.3	25.02.2016	Olivier Frézot (BRGM)	Info on integration of the GI-suite Brokering Framework and SOS Publishing service
1.4	26.02.2016	Paolo Mazzetti (CNR-IIA)	Final draft

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

ABBREVIATIONS AND ACRONYMS

Abbreviation / Acronym	Description
API	Application Programming Interface
CA	Consortium Agreement
CIP	Competitiveness and Innovation Framework Programme
DoW	Description of Work
EC	European Commission
EC-GA	European Commission Grant Agreement
EO	Earth Observation
GI	Geographic Information
ICT	Information and Communication Technology
PSP	Policy Support Programme
RM-ODP	Reference Model for Object Distributed Processing
TL	Task Leader
VH	Virtual Hub
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*. Most of the concepts in this apparently clear statement of objectives are controversial and need a clarification. 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 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, ENERGIC OD adopts an agile methodology allowing rapid development in response to new requirements. It will have three main yearly iterations with fixed objectives for demonstration in reviews and events.

ENERGIC OD will deploy one regional VH in the Berlin metropolitan area and 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 seventh European-level 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 made possible on local infrastructures, possibly managed by one or more ENERGIC OD partners, or on private and 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. A first version of the document was 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.

The present version aims to respond to a recommendation by the reviewers included in the Technical Review meeting report: "Update deliverable to be submitted, incorporating the knowledge coming from WP2, WP3, WP4 and in articulation with WP6 in order to incorporate specific requirements from applications."

This document is the System Definition Document as described in the IEEE Guide to the Software Engineering Body of Knowledge, aiming at "*lists the system requirements along with background information about the overall objectives for the system, its target environment, and a statement of the constraints, assumptions, and non-functional requirements*" [1].

Although the development phase will be carried out inside the Consortium, therefore without the need to establish "*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.

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

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

The third section 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 describes the ENERGIC OD 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 fifth section introduces the agile development approach that is adopted by the ENERGIC OD project, and the sixth and final section reports the deployment plan and achievements at project-month 16.

The Annex I summarizes the main changes between the first and the second version of the present deliverable.

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 <u>development of virtual hubs</u> that <u>facilitate the use of open (freely available) geographic data</u> 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" [2]. 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" [3].

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 which are specific for different disciplinary domains, such as HDF, netCDF and GRIB for EO data, ESRI Shapefile or OGC GML for feature type information. Proprietary formats are still widespread;
- *Data providers* are mainly focused on data sharing services. As for data models and formats, several standards have been designed and adopted in different disciplinary domains. For example, in the biodiversity context TDWG standards are widely adopted, in the meteo-ocean community THREDDS Data Server is a widespread technology. OGC standard services are commonly adopted in the GIS community. Light specifications like KML (now an OGC standard) or OpenSearch are also common. OAI-PMH is a standard for long-term preservation archives.
- *Overarching initiatives* influence technological aspects in several ways, in particular on data management (e.g. 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 this issue 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 [4]. 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 [5].

Concerning ENERGIC OD, the call provides few hints limiting the scope to "*open (freely available) geographic data*" [6]. 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 such as GMES data)". 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 under certain conditions to the data produced by other satellite missions referred to as 'Contributing Missions'" [7]. 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 (Open Data Survey) activities, ENERGIC OD has decided to adopt an operational definition for Open Data [8] 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 freely and generally accessible but not exclusively free of charge

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 and adapted 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" [9]. A software architecture adopting the (message) broker pattern is usually referred as a hub-and-spoke architecture [10].

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 ENERGIC OD proposal built on those definitions and requirements, saying that an 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) the use of Virtual Hub resources, and finally, c) the use of applications based on Virtual Hubs. They are

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 End User	VH End User	The VH End User is a person who accesses the VH through the VH portal or 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.
Virtual Hub App Developer	VH App Developer	The VH App Developer is a person who develops and manages applications based on the VH APIs.
Virtual Hub Consumer	VH Consumer	A VH Consumer is a person who makes use of VH capabilities, which is either a VH End User or a VH App Developer (or both).

Table 1 Description of Virtual Hubs actors

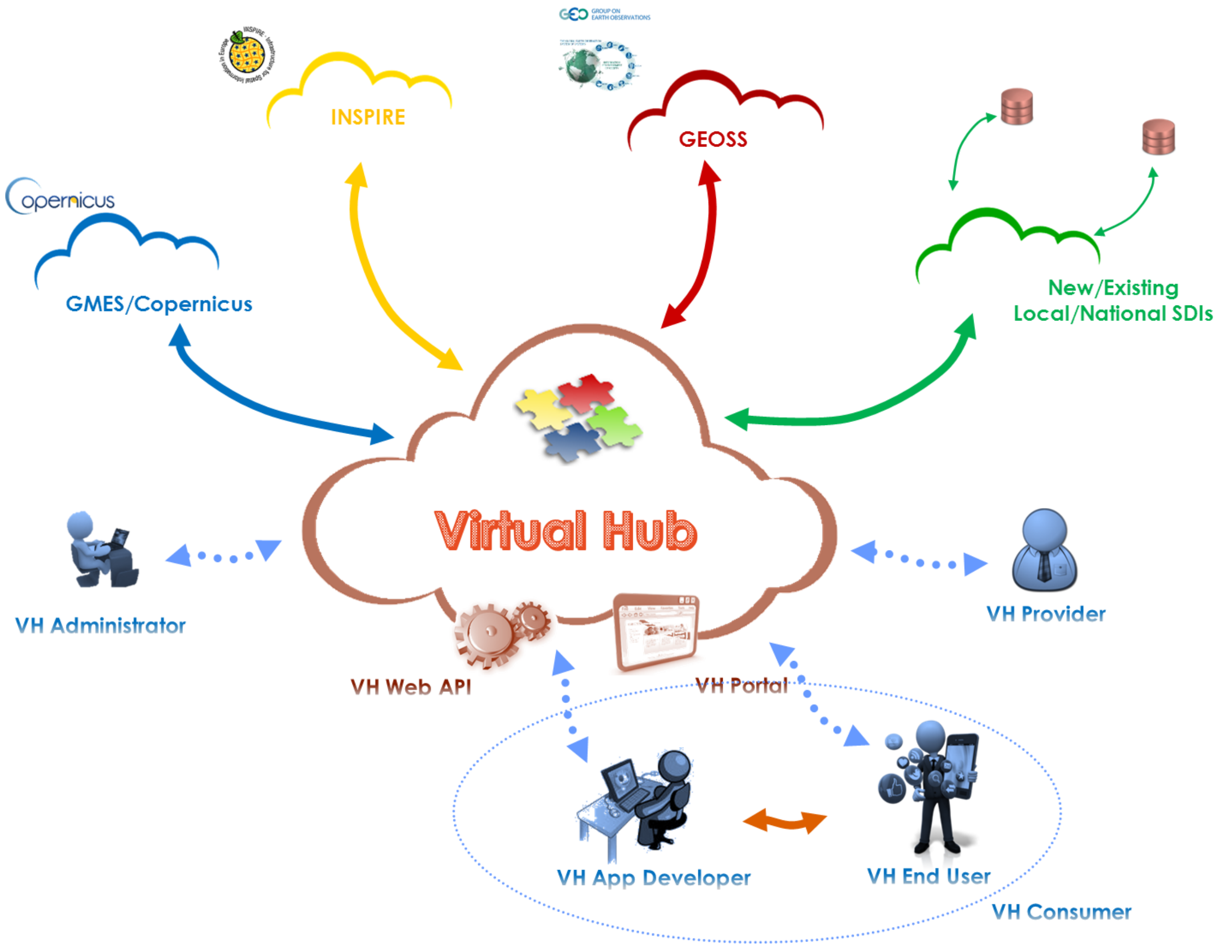

Figure 3 Virtual Hubs actors (in blue)

3.2 User requirements

In the ENERGIC OD work plan, a specific Work Package (WP4) ending at M16, was dedicated to "Requirements and specifications: SDI, data harmonisation and applications addressing user needs" coordinating activities on the definition of user needs. Finally, ENERGIC OD User Requirements are collected from different sources:

a) Call text [6]
b) ENERGIC OD DoW [11]
c) Elicitation of user requirements from ENERGIC OD applications in WP6 "Development of new innovative applications" [12]
d) Elicitation of general user needs in WP4 "Requirements and specifications: SDI, data harmonisation and applications addressing user needs" [13]
e) Previous work in relevant initiatives and programmes at national, regional, European and international level (including Copernicus, INSPIRE, GEOSS)

In terms of user requirements, 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. The high-level required use-cases are those required to support the typical data sharing scenario shown in Figure 4, including Publishing (supporting upload of relevant resources), Discovery (supporting search for relevant resources), Evaluation (supporting inspection of resources to evaluate 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 dataset usage may be a data product to be published. The figure also shows a Management use case which encompasses all the information life-cycle.

Due to the focus of ENERGIC OD (facilitate the use of geospatial open data) and the need of sharing heterogeneous resources within the project and with the outside world, a specific attention on interoperability issues is required.

Figure 4 The typical high-level scenario in the geospatial domain

3.3 Constraints and assumptions

In keeping with its main objective of facilitating the use of geospatial open data, ENERGIC OD tries to pose minimal constraints and assumptions. They are listed in Table 2.

Code	Name	Description
C1	Data discovery service	Each data source MUST make metadata for each dataset accessible through the Web. It may happen through a full catalogue service, or an inventory as simple as a HTTP download of a metadata file.
C2	Metadata	Each dataset MUST be described with a minimal set of metadata including geospatial coverage, temporal extent,

		title, abstract (description), keywords. It SHOULD be described with a more complete set of metadata such as the INSPIRE Profile of ISO 19115
C3	Data access service	Each data source MUST make datasets accessible through the Web. It may happen through a full access service (including advanced functionalities of subsetting, interpolation, etc. such as OGC WCS), or a download service as simple as a HTTP download.

Table 2 ENERGIC OD constraints and assumptions

3.4 System Requirements

ENERGIC OD System Requirements are collected from different sources:

a) Call text [6]
b) ENERGIC OD DoW [11]
c) Elicitation from user requirements (see Section §3)
d) Specific requirements analyzed in in WP4 "Requirements and specifications: SDI, data harmonisation and applications addressing user needs" [14]
e) Specific requirements from ENERGIC OD applications in WP6 "Development of new innovative applications" [12]

Table 3 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 at least the following protocols to communicate with data sources: a) OGC CSW 2.0 (ISO profile) b) OGC WCS (getCapabilities operation), c) OGC WFS (getCapabilities operation), d) OGC WMS (getCapabilities operation), e) OGC WMTS (getCapabilities operation), f) SPARQL, g) FTP h) Socrata Open Data
FR1.2	Dataset discovery protocols (clients)	The system supports at least the following protocols to communicate with clients:

		a) OGC CSW 2.0 ISO Profile b) OpenSearch
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 connect to RDF knowledge bases published with SPARQL/SKOS interface.
FR2.2	Semantic discovery – knowledge bases	The system is able to access GEMET (GEneral Multilingual Environmental Thesaurus) thesaurus for supporting multilingual discovery. Other knowledge bases of potential interest are AGROVOC (Multilingual agricultural thesaurus), UNESCO Thesaurus, NUTS (Nomenclature of Territorial Units for Statistics), CaLAThe (Cadastre and Land Administration Thesaurus, version 1.0), EUROVOC v4.3, GEMET-INSPIRE themes, INSPIRE-Feature Concept Dictionary, INSPIRE-Glossary, ISO-19119 geographic services taxonomy
FR3	Dataset access	The system provides access to datasets from heterogeneous data provision systems
FR3.1	Dataset access protocols (data sources)	The system supports at least the following access/visualisation protocols to communicate with data sources: a) OGC WCS, b) OGC WFS, c) OGC WMS, d) OGC WMTS
FR3.2	Dataset access protocols (clients)	The system supports at least the following access/visualisation protocols to communicate with clients: a) OGC WCS, b) OGC WFS, c) OGC WMS, d) OGC WMTS
FR3.3	Dataset access formats (data sources)	The system supports at least the following data formats, at least through one access protocol (see FR3.1) to communicate with data sources: a) KML, b) GML, c) SHP, d) netCDF, e) JPG f) JPEG2000 g) TIFF

		h) GeoTIFF
FR3.4	Dataset access formats (clients)	The system supports at least the following data formats, at least through one access protocol (see FR3.2) to communicate with clients (possibly through on-the-fly transformation, see FR4): a) KML, b) GML, c) GeoJSON, d) netCDF, e) JPG f) JPEG2000 g) GeoTIFF
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.).
FR4.1	Dataset transformation protocols and CRS	The system supports at least: a) Specific external transformation services exposed with an OGC WPS interface, b) Conversion from WFS/GML into GeoJSON, c) Conversion form SHP into GeoJSON, d) CRS transformation from any CRS supported by INSPIRE-compliant systems to EPSG:4326
FR5	Support of legacy/proprietary formats	The system must be able to support a set of relevant legacy/proprietary data formats. This is necessary to support relevant open data source with low rate of openness (see §2.3).
FR5.1	Support of legacy/proprietary formats	The system supports at least: a) SHP
FR6	Data processing	The system must be able to process data. Only general processing, (i.e. processing required by multiple applications) will be supported by the system. Specific processing, tailored to single applications is up to the application developer.
FR7	AAA	The system must support Authentication, Authorization

		and Accounting allowing collecting information about the use for both technical and marketing purposes.
FR8	Data Publishing	The system must support data publishing from data collection applications.
FR8.1	Data publishing protocols	The system will support upload of observation using OGC SOS protocol.
NFR1	Seamless discovery and access	The system provides discovery (FR1) and access (FR3) of heterogeneous data sources (FR1.1, FR3.1) through any of multiple standard interfaces (FR1.2, FR3.2).
NFR2	Multiple distribution strategies	The system can be replicated on multiple instances depending on technological or marketing choices. ENERGIC OD will demonstrate this flexibility deploying: a) 1 local-level Virtual Hub in the Berlin region b) 5 national-level Virtual Hubs in France, Germany, Italy, Poland, Spain
NFR3	Multiple deployment strategies	The system can be deployed according to different strategies: a) local infrastructures b) private Infrastructure-as-a-Service (IaaS) clouds (virtualised local infrastructures) c) 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)
NFR4.1	APIs implementation	The system supports at least: a) server-side open interface b) Web APIs (HTML5-JavaScript-CSS library)
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
NFR11	Accuracy	The system should not introduce loss of data quality (e.g. in data transformations)

Table 3 ENERGIC OD system requirements

4 ENERGIC OD ARCHITECTURAL PRINCIPLES

4.1 Open Software Architectures

Taking into account the main constraint of the ICT PSP: "The ICT PSP does not support research activities; 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.

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 4). A first classification [15] 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	When: 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
Who	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 4 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

Interoperability is recognized as the main challenge for ENERGIC OD. As the call says: "Solutions should lead to an easier discoverability and use of geographic information available for use in innovative applications and services, and where possible draw together datasets from different sources". 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* [16]. 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.

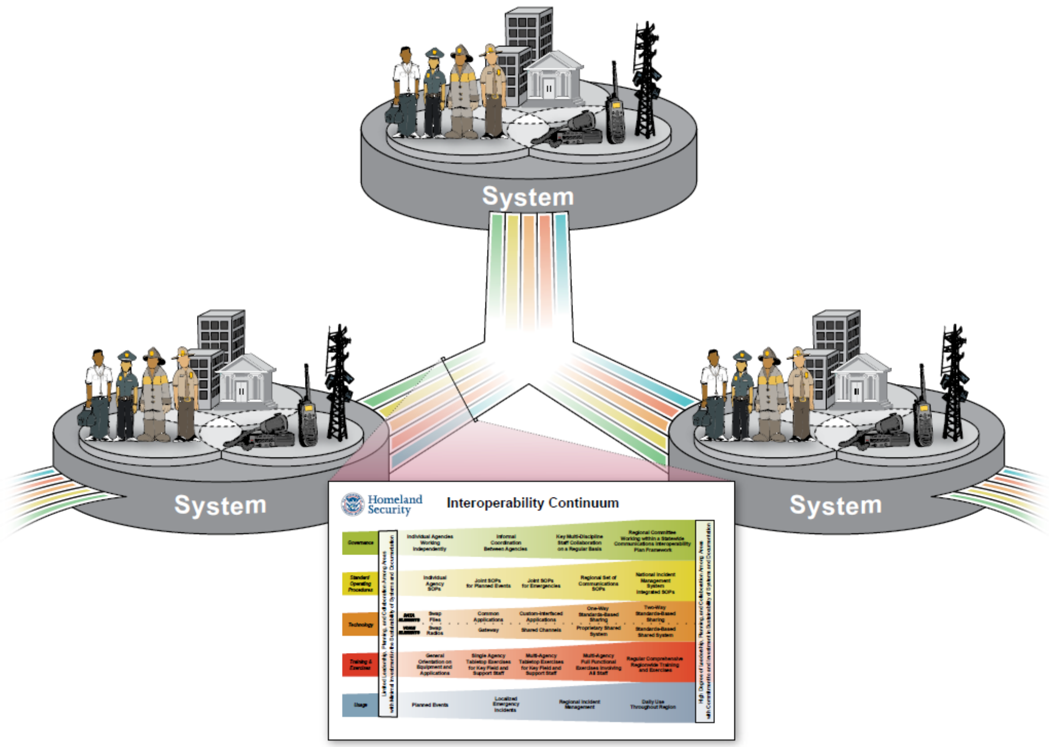

Figure 5 System of Systems in Practice – from [17]

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* [18], 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" [19]. 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" [19]. 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 noteworthy, that the former is currently addressed by the Brokering Governance WG[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+brokering approach. Standardization helps to reduce the redundant heterogeneity, while brokering addresses the remaining irreducible heterogeneity.

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

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.

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

[4] http://www.tdwg.org

[5] www.unidata.ucar.edu/software/thredds/current/tds/

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

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: 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).

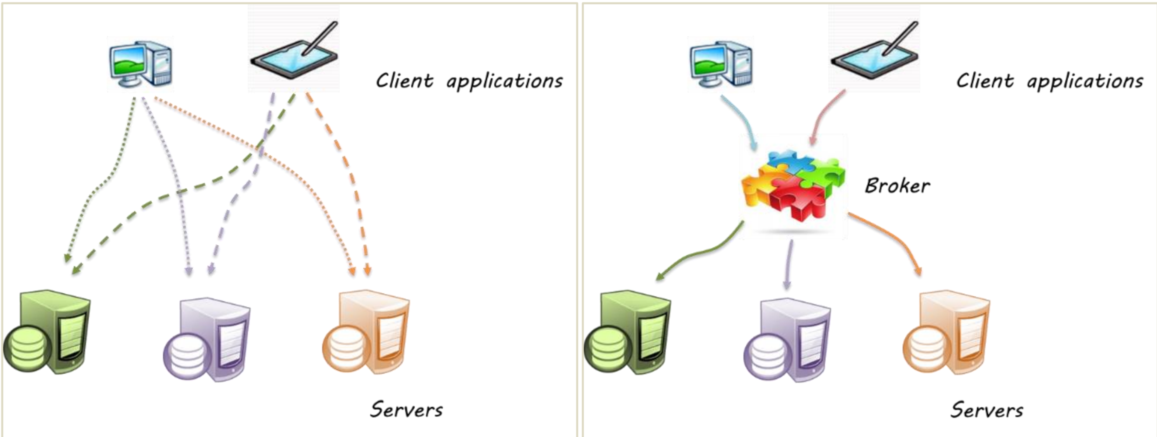

Figure 6 Client-Server vs. Brokered Architectures

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*" [20].

The cloud model includes three different kinds of services [20]:

- *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 may also 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" [21]. 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" [22].

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*" [22].

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 (RM-ODP) [23]:

- 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;*
> *[23]*

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 documented in D6.1 "Application based requirements and standards catalogue" [12]. Further investigations of enterprise aspects will be the objective of WP8 on Exploitation.

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.*
> *[23]*

Figure 7 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).

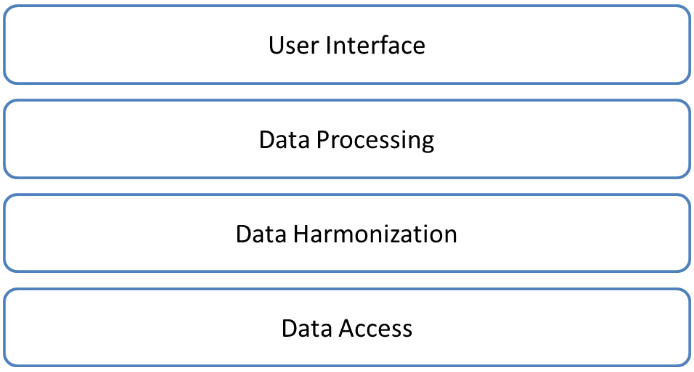

Figure 7 ENERGIC OD layered architecture

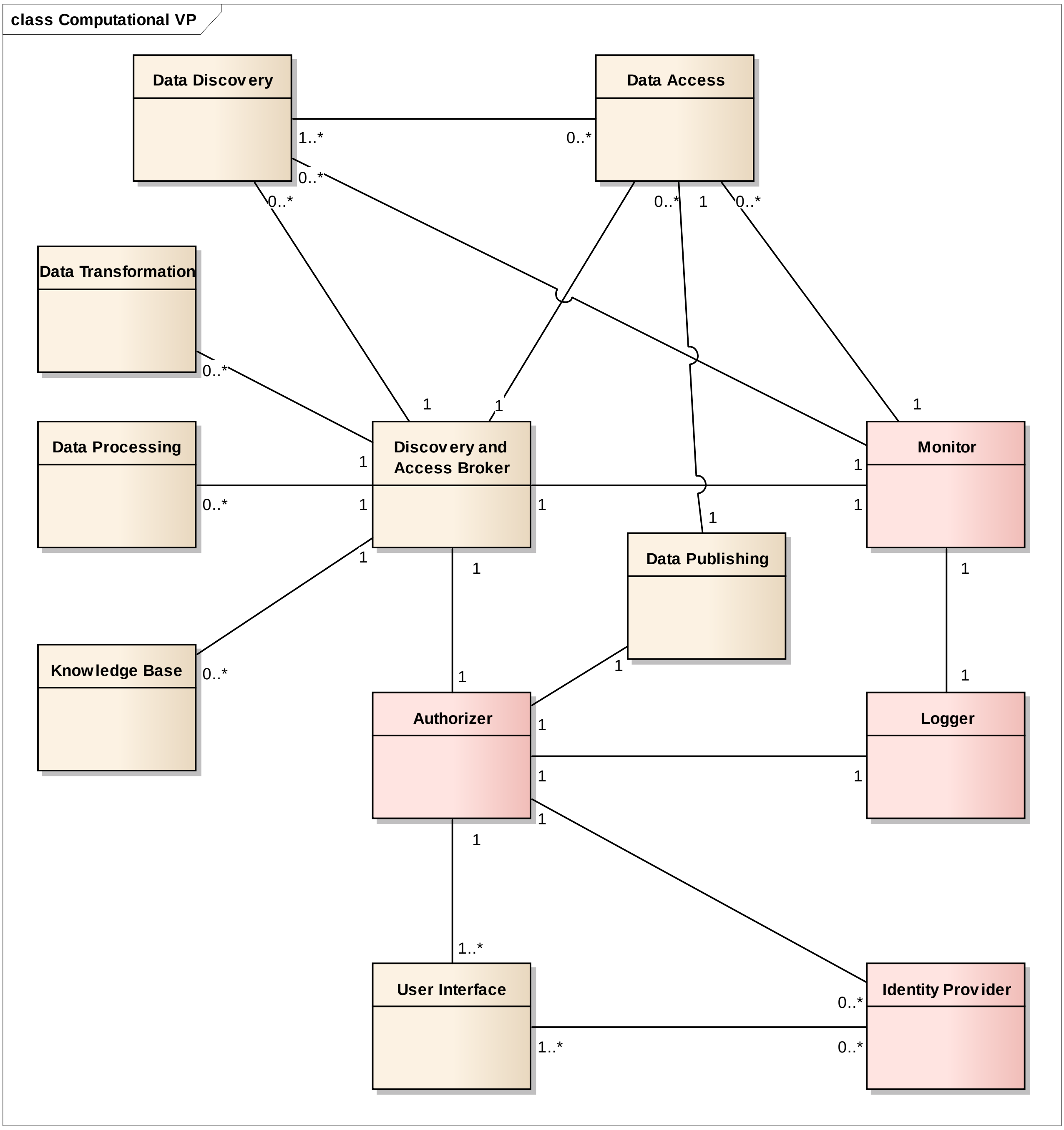

Figure 8 Main components of the ENERGIC OD architecture

Figure 8 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. facilitating 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 5 along with a reference to the functional and non-functional requirements they contribute to fulfil.

Component	Description	Relevant requirements/constraints
Data Discovery	It provides discovery of open data from specific data sources	C1, C2 FR1, FR2
Data Access	It provides access to open data from specific data sources	C3 FR3
Discovery and Access Broker	It accesses Data Discovery and Data Access components providing harmonized access to heterogeneous data sources.	FR1, FR2, FR3, FR5, FR7 NFR1, NFR10
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.	FR7
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	FR7 NFR8
Monitor	It checks availability and status of data sources	NFR5
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 NFR5, 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 5 ENERGIC OD main components

From Table 5 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). As a logical architecture, it addresses all the functional requirements also impacting on some non-functional requirements. Non-functional requirements are addressed by the distribution architecture discussed in the Engineering Viewpoint in section 5.5 (e.g. NFR6, NFR7), and by the implementation and deployment choices described in sections 6 and 0 (e.g. NFR2, NFR3, NFR4 and NFR11).

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. [23]

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 annotated 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", WP4 on "Requirements and specifications: SDI, data harmonisation and applications addressing user needs" and WP6 "Development of new innovative applications" 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 6 and Table 7 show a list of potential Virtual Hub protocols beyond those strictly required, respectively for connected data source and exposed interfaces. They are collected from previous activities on geospatial data systems of systems and from the ENERGIC OD activities in the mentioned work packages.

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 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
Socrata Open Data API	Data discovery service
ESRI shapefile	File format
.KML	File format

CityGML GML	File format
JPG	File format
GeoJSON	File format
TIFF	File format

Table 6 List of relevant protocols supported by data sources in the geospatial domain

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 7 List of protocols supported by the Virtual Hubs exposed 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 [24]. 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 9 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 includes either the proposed keyword or any of its supported translations. This is extremely important whenever there is not any obligation to compile metadata in a specific language.

5.5 Engineering Viewpoint

Engineering VP is concerned with the infrastructure required to support system distribution.
[23]

Figure 9 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.

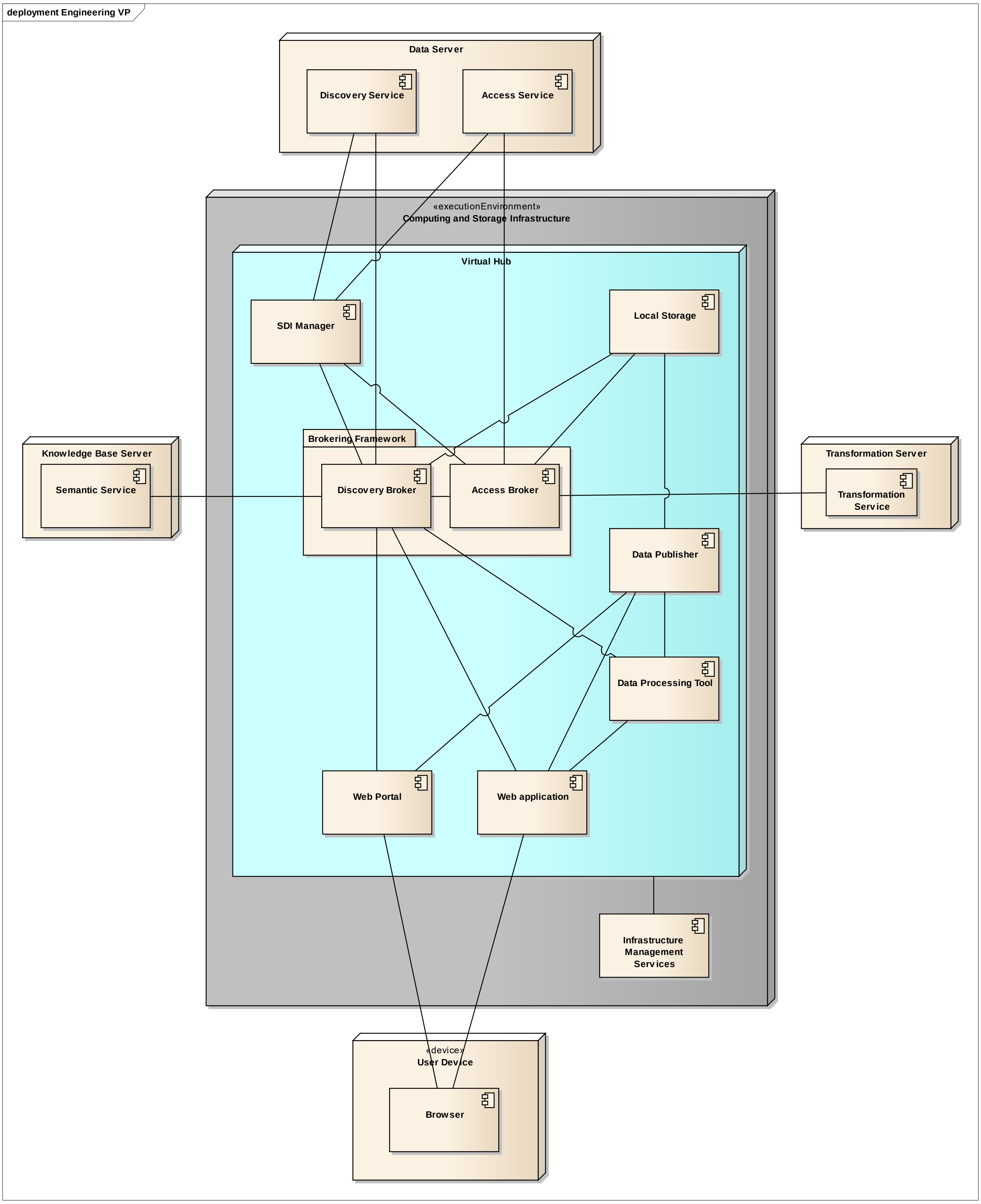

Figure 9 Engineering view of the ENERGIC OD Virtual Hub architecture

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. According to the brokering approach, 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 five Virtual Hubs at national level and one at local 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 private or public cloud offering Infrastructure-as-a-Service capabilities.
Knowledge Base Server	A Knowledge Base Server is a node providing services accessible by the Virtual Hub for semantic enhancements. According to the brokering approach, no assumption is made about communication protocols.
Transformation Server	A Transformation Server is a node providing services accessible by the Virtual Hub for data transformations (e.g. re-projection on different Coordinate Reference Systems, format encoding, sub-setting, change of resolution and interpolation) enhancements. According to the brokering approach, no assumption is made about communication protocols.
User Device	A User Device is a node hosting user's applications. It can be a desktop, or a mobile device. The only assumption is that it is able to host a (modern) Web browser.

Such 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 10 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.

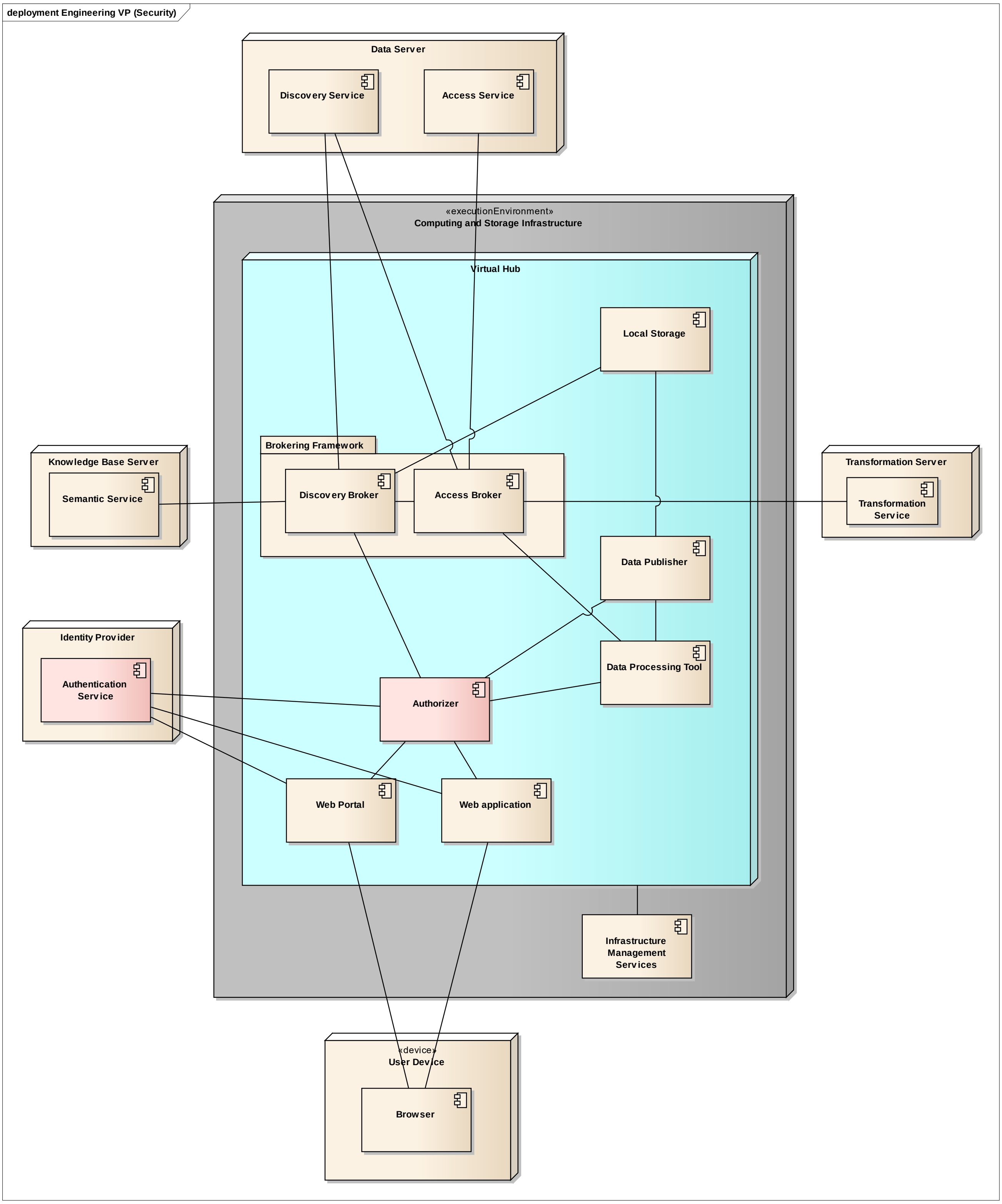

Figure 10 Engineering view of the ENERGIC OD architecture showing the main security components

5.6 Technology Viewpoint

Technology VP is concerned with the choice of technology to support system distribution.
[23]

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. WP2 on "State of the art in R&D, projects & technologies" carried 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 [25]. They are analysed below, together with the ones proposed by the partners in the DoW, while Table 8 shows the final approach and finally selected solutions for each of the main architectural components.

5.6.1 Brokering Framework

The main technological choice regarding the architecture and implementation of the Virtual Hubs is the selection of the brokering framework. The DoW proposed two frameworks developed and maintained by ENERGIC OD partners: the GI-suite Brokering Framework and the GIS-Broker, while WP2 identified through the SWOT review of certain FP7 projects (namely, NETMAR, ENVISION, Apps4EU, ENVIROFI, FINODEX and EUBrazilOpenBio) some others: CSWM 1.0, Semantic Catalogue, DataTank, Datalift, Environmental Georeferenced Observation Service and gCube.

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

The suite is composed of the following components:

- GI-cat: a discovery broker;
- GI-sem: a semantic broker;
- GI-axe: an access broker;
- GI-quality: a quality broker;
- GI-BP: a business process broker.
- GI-go: a thick client to test the suite;
- 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.

The review of the SWOT analyses in WP2 revealed that the GI-suite Brokering Framework has already being used in several projects and has been improved through them (EuroGEOSS, ENVIROFI, GeoViQua and 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 ENERGIC OD 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 ENERGIC OD.)

GIS-BROKER from SRP is a Java based framework (J2EE) for cross platform integration of data storage systems (GIS, RDBMS), applications and web services within geospatial data aided business processes and special proceedings.

It is a solution for building SDIs to share GIS data (feature data types, raster images, etc.) with advanced functionalities including:

- Catalogue (OGC CSW compliant; management of metadata)
- Map Management
- Feature Management
- GIS Processing
- GIS-Data Management
- Service Relation Management
- User & Resource Management
- System Management
- Access Services and System Adapters
- Web Services Interfaces for easy integration into existing IT-environments
- GeoPortal & Application Framework
- Client Modules
- Communication Interfaces

GIS-BROKER-Framework is a mature and proven technology for the creation of infrastructure nodes, in real use in several public administrations:

- Basic technology of the Spatial Data Infrastructure of Berlin
- GeoPortal of Berlin "FIS-Broker"
- Real Estate Information System of the Liegenschaftsfonds GmbH of Berlin
- GeoInformation System of several administrative units in Berlin (Urban planning authorities, land surveying offices,…)

NETMAR Catalogue Services for the Web Mediated (**CSWM 1.0,** an OGC CSW semantically enabled) and the ENVISION **Semantic Catalogue** for enhanced access to OGC based service offer lower levels of functionality than the one of the GI-suite (as they only deal with data search but not with data access) and their level of maturity is difficult to evaluate, as both are result of a single project and no information on further developments or updates have been found.

For a small-scale brokered approach, the Apps4EU project recommended the use of Open Data with the DataTank platform (for a non-semantic API approach) and the Datalift platform (semantic API with a Sesame backend). **DataTank**[7] can provide data in CSV, XML, JSON, SHP and KML [26]. The **Datalift**[8] platform can use content produced by the DataTank and relational databases and publish it enriched as Linked Data [26]. However, the support of geographic information in these two platforms is ancillary: DataTank does not support CRS and the query language does not support spatial queries; Datalift does not support geographic information yet.

FIWARE is a platform seeking to provide an open, public and royalty-free architecture and a set of open

[7] http://thedatatank.com/

[8] http://datalift.org/

specifications to allow developers, service providers, enterprises and other organizations to develop products that satisfy their needs while still being open and innovative. FIWARE provides a rich library of components offering a number of added value functions offered “as a Service”. These components, called “Generic Enablers” (GEs), provide open standard APIs that make it easier to connect to the Internet of Things, process data and media in real-time at large scale, perform Big Data analysis, manage security and access control or incorporate advanced features to interaction with the user. ENVIROFI **Environmental Georeferenced Observation Service**[9] acts like a FIWARE geo-broker, although its capabilities are limited to access to OGC WFS and SOS services, shapefiles and a couple of proprietary RESTful web services and only can be accessed as HTTP REST queries or through the OGC SOS interface. The only advantage over the GI-suite Brokering Framework would be its integration in the FIWARE context. However, it is noteworthy that the Discovery Broker (GI-cat) and Access Broker (GI-axe) components of the GI-suite are environmental Specific Enabler of the Future Internet Public Private Partnership, accessible through the ENVIROFI Catalogue[10]. (ENVIROFI proposed to move some of its Specific Enablers to FIWARE but the choice of FIWARE was to not consider geospatial functionalities as requirements of the Future Internet core platform.)

The **gCube** framework[11] can act as a broker. GCube is a large software framework dedicated to scientists and designed to abstract over a variety of technologies belonging data, process and resource management on top of Grid/Cloud enabled middleware. The gCube system provides support for: OGC standard protocols like WFS, WCS, WMS, WPS, OAI providers for document data, SPARQL endpoints and Biodiversity data sources, including OBIS, GBIF and Catalogue of Life. GCube seems mature enough (7 years of history and currently in version 3), is supported by a large community, although the frequency of the updates to the source code seems to have been decreasing and its concept is too focused on the execution of scientific workflows. The number of access and exposed interfaces is high, but less than the ones offered by the GI-suite Brokering Framework. Nevertheless, if considered appropriate by the exploitation plans to be defined in WP8, it could be a choice for the implementation of scientific, Big Data intensive VHs, and could be easily wrapped by the GI-suite Brokering Framework in order to provide all VHs with a common, consistent access interface.

The choice of the GI-suite Brokering Framework as the central brokering component of the VHs is determined firstly by its features. Its broker capacities outrun the ones of the other frameworks analysed in WP2, as it has been discussed above. 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 implantation 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 ENERGIC OD 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 WP4 and the WP6 pilots applications lays within the ENERGIC OD consortium.

Additionally, the GIS-Broker will be used too as part of the brokering infrastructure. Its wide adoption in the Berlin-Brandenburg area and its capacity to federate SDIs for GIS data allow for its use as a subnational VH, that would be connected by the GI-suite Brokering Framework in order to provide all VHs with a common, consistent access interface. The GIS-Broker is also of proven maturity, being used for the creation of infrastructure nodes, in real use in several public administrations. It is also provided by one of the partners, SRP, so its development is not dependant on third parties outside the ENERGIC OD consortium.

[9] http://catalogue.envirofi.eu/enablers/environmental-georeferenced-observation-service

[10] http://catalogue.envirofi.eu/enablers?field_enabler_category_tid=10

[11] http://www.gcube-system.org/

5.6.2 Semantic Service

WP2 identified a set of controlled vocabularies as useful for the Virtual Hubs: GEMET[12] (GEneral Multilingual Environmental Thesaurus), AGROVOC[13] (Multilingual agricultural thesaurus), UNESCO Thesaurus[14] and NUTS[15] (Nomenclature of Territorial Units for Statistics).

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

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

The semantic service published by EC-JRC and providing a set of aligned thesauri will be initially used for multilingualism, suggestions, and semantic queries. The GI-suite Brokering Framework is already able to access the EC-JRC Semantic service through its GI-sem semantic module.

[12] https://www.eionet.europa.eu/gemet/
[13] http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus
[14] http://databases.unesco.org/thesaurus/
[15] http://ec.europa.eu/eurostat/web/nuts/overview

Whenever required, other knowledge base can be developed and published using open source tools supporting SPARQL/SKOS. WP2 SWOT analyses of five projects [25] provided detailed information on the technologies and tools they used regarding Linked Data (structured and semantically interlinked data), based on W3C specifications like the RDF family of languages to represent ontologies (mainly OWL, SKOS and XKOS) and the SPARQL Protocol and RDF Query Language. However, a long list of tools supporting these specifications was used, due to the fact that Linked Data is an issue under research, with many, different technological approaches. However, technologies developed under the Apache Software Foundation (Clerezza[16], Stanbol[17], Solr[18], Marmotta[19], Jena[20], D2RQ[21]) were prevalent.

As the requirements established by WP4 and the WP6 pilots applications do not include access to Open Data in Linked Data form, as this format it is not currently widely used due to its novelty and the management complexity it poses, no technological decision regarding implementing tools is needed.

The need of access to Linked Data in the ENERGIC OD VH will be limited to the access of external knowledge bases that will be used to semantically enrich queries. As the GI-suite Brokering Framework already supports access to semantic services through SPARQL/SKOS, the software and tools used by these external tools is irrelevant, provided that they can be queried through SPARQL/SKOS.

5.6.3 Transformation Service

WP2 and WP4 identified coordinate transformation as a core service to be provided by the VHs within its brokering capabilities. However, there is not any official, internationally agreed service interface specification available for coordinate transformations. The approach adopted by OGC is to incorporate the only mandatory operation Transform defined in of the OGC Discussion Paper 'Web Coordinate Transformation Service' (WCTS) as an Application Profile of the OGC Implementation Specification 'Web Processing Service' (WPS). The OGC's WCTS specification is available as an OGC Discussion Paper (version 0.4.0). There seems to be no activity in the OGC to advance the WCTS specification to official level.

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

WP2 SWOT analyses identified **Proj4js**[22] as a JavaScript library suitable to transform point coordinates from one coordinate system to another, including datum transformations, that could be used by some applications. However, the preferred approach for coordinate transformations required by applications would be to rely on the VHs.

5.6.4 Local Storage

VHs, as broker-based hubs, mediate among data sources and data users, with no requirement of storing the mediated data. However, a VH local storage may be needed to support crowdsourcing, and Volunteered Geographic Information (VGI) providers. Actually, the GI-suite Brokering Framework currently uses a XML database (either open-source eXistdb[23] database or MarkLogic[24] enterprise NoSQL database for higher

[16] http://clerezza.apache.org/
[17] http://stanbol.apache.org
[18] http://lucene.apache.org/solr/
[19] http://marmotta.apache.org/
[20] https://jena.apache.org/
[21] http://d2rq.org/
[22] http://proj4js.org/
[23] http://exist-db.org/
[24] http://www.marklogic.com/

performances) for storing harvested metadata and configuration. In principle, it could be used to store new datasets also, however a dedicated tool with specific functionalities for uploading geospatial datasets may be preferred. WP2 identified a set of geographical databases. Although different applications will have different local storage needs, from the WP2 analysis it is possible to make the general suggestion of using PostGIS/PostgreSQL.

PostGIS[25] is an open source software program that adds support for geographic objects to the **PostgreSQL**[26] object-relational database. PostGIS implements the OGC Simple Features for SQL specification.

MySQL[27] is an open-source relational database management system that implements spatial extensions as a subset of the SQL with Geometry Types environment. However, MySQL does not implement some features of OGC Simple Features.

Oracle Spatial[28], now renamed as Oracle Spatial and Graph, forms a separately-licensed option component of the Oracle Database. The spatial features in Oracle Spatial and Graph aid users in managing geographic and location-data in a native type within an Oracle database. The geospatial feature of Oracle Spatial and Graph provides a SQL schema and functions that facilitate the storage, retrieval, update, and query of collections of spatial features in an Oracle database.

Esri ArcSDE[29] (Spatial Database Engine) is a server-software sub-system produced and marketed by Esri that aims to enable the usage of Relational Database Management Systems for spatial data. The spatial data may then be used as part of a geodatabase.

Strabon[30] is a semantic geospatial and temporal DBMS for storing and querying geospatial data that changes over time. Strabon is an implementation of the data model stRDF, the query language stSPARQL and the respective part of the OGC standard GeoSPARQL. stRDF and stSPARQL extend RDF and SPARQL 1.1 respectively providing a function set and data types for making the querying of spatiotemporal information via stSPARQL or GeoSPARQL possible, such as finding spatial and temporal relations between two resources. Strabon is an open source application dating 2010, currently in version v3.2.10.

5.6.5 Data Publisher

VHs data publishing requirements are limited to the datasets uploaded to a VH local storage in order to support specific applications, such as crowdsourcing and Volunteered Geographic Information (VGI). The GI-suite Brokering Framework can see the local storage as one of the many data sources, offering any of its supported interfaces to access these data.

The analysis of ENERGIC OD pilot applications in WP6 shows that the applications requiring publishing consider the data publisher outside the boundaries of the VH. This is reasonable since the pilot applications assume that a publishing system may already exists outside the VH, and then it is connected to the VH for data retrieval. However, it is noteworthy that pilot applications aim to validate the VH approach, and several other potential applications may indeed require to publish data without knowing of existing external services. Therefore, the solutions implemented for data publishing during the pilot app development may be reused as internal VH components to support those potential crowdsourcing and VGI applications. Four pilot applications explicitly mention data publishing: a) A3: Coastline Evolution Monitoring by BRGM develops an OGC SOS data publishing service; b) A4: OnoMaP! by CNRS-IRSTV uploads data on a RDBMS; c) A9: geoDEMOS

[25] http://postgis.net/
[26] http://www.postgresql.org/
[27] http://www.mysql.com/
[28] http://www.oracle.com/us/products/database/options/spatial/overview/index.html
[29] http://www.esri.com/software/arcgis/arcsde
[30] http://www.strabon.di.uoa.gr/

by SRP allows uploading and annotation of GIS datasets through the GIS-Broker tool; d) A10: Sensor Open Data Portal by ALKANTE allows sensor data upload. Therefore, BRGM, CNRS-IRSTV, SRP and ALKANTE may become also VH technology providers for the Data Publisher. The DoW of ENERGIC OD and the WP2 identified several frameworks and tools that can be useful for this purpose: SpatiumCube, GeoServer, Degree, PostgreSQL, EasySDI, CKAN, GeoBatch, MapServer, THREDDS, Constellation, Dapper. Additionally, CatMDEdit and EUOSME enable metadata management and Sync'Serv enables database synchronization.

5.6.6 Data Processing Tool

The objective of a Virtual Hub is to facilitate the use of open geospatial data. As such, an ENERGIC OD VH does not focus on processing. However, processing functionalities, which are useful for many potential applications are good candidate to be provided by the VH. WP2 identified potentially useful frameworks and tools.

Web Processing Service[31] (WPS) is an OGC standard that provides a standard interface that simplifies simple or complex computational processing services accessible via web services. The WPS standard defines how a client can request the execution of a process, and how the output from the process is handled. WP6 applications could make use of WPS interface to encapsulate complex data processing in order to be reused in other contexts. **PyWPS**[32] (Python Web Processing Service) consists of several WPS wrappers for GRASS[33] functionality. Although the maturity cannot be assessed, GRASS is a free and open-source Geographic Information System (GIS) software suite used for geospatial data management and analysis, image processing, graphics and maps production, spatial and temporal modelling, and visualization. GRASS has been under development since 1982, with the latest stable release (version 7) in 2015. It is a very mature technology to take into account when topological transformations on geographical data are needed.

Other tools and software related to data processing and model chaining were identified by the project DoW and the WP2 SWOT analyses, like **OpenRoute**[34], the **Geospatial Data Abstraction Library**[35] (GDAL), **Taverna**[36] and the **Nasa World Wind Java SDK**[37], that could be of use when implementing data processing chains in the WP6 pilot applications that require them.

5.6.7 Web portals and web applications

A VH should be accessible by both end-users, e.g. through a portal for human-to-machine interaction, and by intermediate users like developers, e.g. through APIs for machine-to-machine interaction. The GI-suite Brokering Framework already provides a test portal (GI-Portal) and web APIs (GI APIs) offering access to the brokering capabilities. They can be used and extended to answer the requests of VH managers and apps.

WP2 identified several useful frameworks and tools, like: **GIS Portal** (from AED-SICAD, a framework for developing web-based GIS applications, mature and adopted in operational settings), **Atl@nte platform** (composed of various open source tools and provided by POLIMI, also a framework for web GIS applications development and mature enough), **GeoExt**[38] (open source JavaScript library for application development, mature and widely adopted technology), **GeoWebCache**[39] (Java web application used to cache map tiles

[31] http://www.opengeospatial.org/standards/wps
[32] http://wiki.rsg.pml.ac.uk/pywps/Main_Page
[33] http://en.wikipedia.org/wiki/GRASS_GIS
[34] http://openrouteservice.org/
[35] http://www.gdal.org/
[36] http://www.taverna.org.uk/
[37] http://worldwind.arc.nasa.gov/java/
[38] http://geoext.org/
[39] http://geowebcache.org/

coming from a variety of sources), **Constellation-SDI**[40] (software framework to create a complete Spatial Data Infrastructure, from cataloguing geographic resources to operating a platform of sensors that feeds back information in real time, based in OGC services), **Dapper**[41] (OPeNDAP web server that provides networked access to in-situ and gridded data), **Liferay Portal**[42] (free and open source enterprise portal project), **MapFaces**[43] (library of JSF components for the construction of web based, geospatial applications). **OpenLayers**[44] (open source JavaScript library for displaying map data in web browsers), **jQuery**[45] (cross-platform JavaScript library designed to make it easier to navigate a document, select DOM elements, create animations, handle events, and develop Ajax applications), **Google Web Toolkit**[46] (GWT, open source set of tools that allows web developers to create and maintain complex JavaScript front-end applications in Java), **Ext JS**[47] (JavaScript application framework for building interactive web applications using techniques such as Ajax, DHTML and DOM scripting), **Web Graphics Library**[48] (WebGL, a JavaScript API for rendering interactive 3D computer graphics and 2D graphics within any compatible web browser without the use of plug-ins), **Cesium**[49] (a JavaScript library for creating 3D globes and 2D maps in a web browser without a plugin), **Simile Timeline**[50] and **Simile Timeplot**[51] (both free and open-source data visualisation web widgets for visualizing temporal data and plotting time series) and the **HTML5 Geolocation API**[52] **(**effort by the W3C to standardize an interface to retrieve the geographical location information for a client-side device)

5.6.8 Authentication Service

WP2 identified several frameworks and tools that can be use as authentication services both by the VHs as by the applications.

OpenID[53] is an open standard and decentralized protocol that allows users to be authenticated by certain co-operating sites using a third party service. This eliminates the need for webmasters to provide their own ad hoc systems and allowing users to consolidate their digital identities. **OAuth**[54] is an open standard for authorization that provides client applications a “secure delegated access” to server resources on behalf of a resource owner. It specifies a process for resource owners to authorize third-party access to their server resources without sharing their credentials. Designed specifically to work with HTTP, OAuth essentially allows access tokens to be issued to third-party clients by an authorization server, with the approval of the resource owner, or end-user. The client then uses the access token to access the protected resources hosted by the resource server. OAuth is commonly used as a way for web surfers to log into third party web sites using their Google, Facebook or Twitter accounts, without worrying about their access credentials being compromised. OAuth is a service that is complementary to, and therefore distinct from, OpenID.

---

40 http://www.constellation-sdi.org/en/
41 http://www.epic.noaa.gov/epic/software/dapper/
42 http://www.liferay.com/
43 http://mapfaces.codehaus.org/
44 http://openlayers.org/
45 https://jquery.com/
46 http://www.gwtproject.org/
47 http://www.sencha.com/products/extjs/
48 https://www.khronos.org/webgl/
49 http://cesiumjs.org/
50 http://www.simile-widgets.org/timeline/
51 http://www.simile-widgets.org/timeplot/
52 http://dev.w3.org/geo/api/spec-source.html
53 http://openid.net/
54 http://oauth.net/

The **Security Assertion Markup Language 2.0**[55] (SAML 2.0) is an OASIS XML-based, open-standard data format for exchanging authentication and authorization data between parties, in particular, between an identity provider and a service provider. SAML dates from 2001; the most recent major update of SAML was published in 2005, but protocol enhancements have steadily been added through additional, optional standards. The single most important requirement that SAML addresses is web browser single sign-on (SSO).

5.6.9 Selected solutions for each of the main architectural components

Table 8 summarizes the technological solutions for the main architectural components in the ENERGIC OD architecture.

Component	Approach	Available solutions
Brokering Framework	Existing + Integration + Enhancement	The **GI-suite Brokering Framework** implements 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, using technologies developed under the **Apache Software Foundation** (Clerezza, Stanbol, Solr, Marmotta, Jena, D2RQ).
Transformation Service	Existing + Development if required	Subsetting is supported through GI-suite internal implementation. Simple interpolation schemes are supported through GI-suite internal implementation Most used Coordinate Reference Systems are supported. Specific CRS transformations will be implemented if needed possibly through access to external services.
Local Storage	Integration	The GI-suite Brokering Framework currently uses **MarkLogic** for metadata storing. Several geospatial open source tools may be used by WP6 pilot applications to implement local storage, like **PostGIS/PostgreSQL** (preferred) or MySQL.
Data Publisher	Integration	**SpatiumCube**, **GeoServer**, **Degree**, **PostgreSQL**, **EasySDI**, and several geospatial open source tools may be used to implement data publishing by WP6 pilot applications. **CatMDEdit** and **EUOSME** enable metadata management to WP6 pilot applications. **Sync'Serv** enables database synchronization to WP6 pilot applications. Solutions to be developed by some ENERGIC OD partners

[55] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security

		applications (Coastline Evolution Monitoring by BRGM, OnoMaP! uploader by CNRS, GIS Broker by SRP, Sensor Open Data Portal uploader by ALKANTE,) may become part of the VH framework.
Data Processing Tool	Integration	**OpenRoute** enables path routing to WP6 pilot applications. WPS interface can be used to encapsulate complex data processing in order to be reused in other contexts by WP6 pilot applications. **Taverna** could be used for more complex workflow manager (if required by WP6 pilot applications).
Web Portal	Development	**GIS-Porta**l, a Stack of Open Source software for geomatics, the **Atl@nte platform**, and several open source tools may be used to implement data portals by WP6 pilot applications.
Web Application	Development (see WP6)	Web APIs (**GI API**) are already available for the GI-suite Brokering Framework. **GeoExt**, **OpenLayers** and several other Javascript libraries are available to develop geospatial Web applications by WP6 pilot 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 8 Technological approach and potential solutions

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 11) [27] 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) [28]:

- 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.

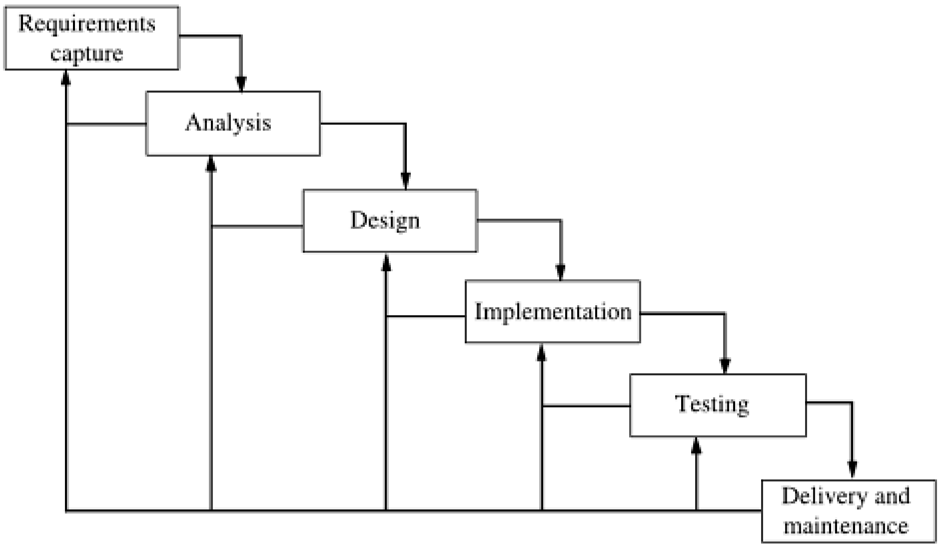

Figure 11 The traditional Waterfall Model (from [27])

Agile methodologies better respond to changes through an iterative process (Figure 12). 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.

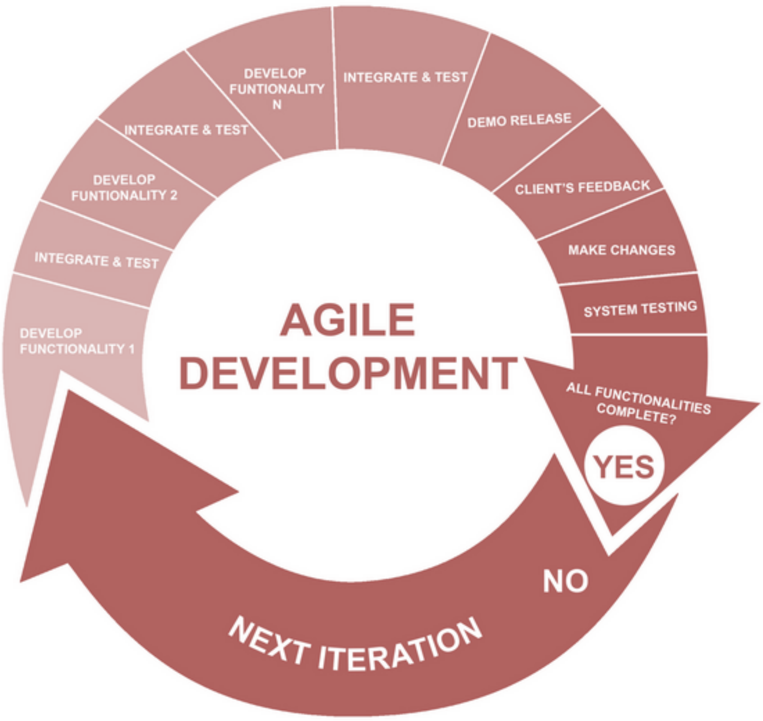

Figure 12 The iterative process in Agile development

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 ended on month 12, releasing the VH Initial Capacity providing discovery and access brokering of multiple data sources.

6.2 System integration

As described in Section 4.1, 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. The different components of the Virtual Hub architecture are then implemented through the integration of selected technological solutions to build a complete framework delivering the requested Virtual Hub functionalities. The first release of the Virtual Hub comprised the GI-suite Brokering Framework. The following releases will include selected components integrated with the GI-suite Brokering Framework to support missing functionalities. During the second year of projects two activities integration are started: a) integration of the GI-suite Brokering Framework with the SRP GIS-Broker for enhanced support of GIS data sources and b) integration of the GI-suite Brokering Framework with a SOS-compliant data publishing system by BRGM. At the moment of the release of this deliverable (February 2016) the second activity was just started. Therefore the following sub-sections will describe the GI-suite Brokering Framework, the integration plan with the GIS-Broker by SRP, and the initial decisions for the integration with the BRGM SOS publishing service.

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 ENERGIC OD are:

- *Discovery broker (GI-cat)*: a component which is able to connect disparate (distributed and heterogeneous) metadata sources, exposing them through a set of standard catalogue interfaces. By means of metadata harmonization and protocol adaptation, it is able to search metadata from different sources and transform query results to a uniform and consistent metadata model. GI-cat mediates among the connected metadata sources interfaces, and harmonizes their metadata mapping them to an internal schema based on ISO 19115 (GI-cat metadata model). Each query request sent through the external interfaces is performed against all the connected sources based on the internal schema. GI-cat supports both distributed queries (for external sources exposing a catalogue service) and harvesting. Harvesting can be adopted for enhance query performances for catalogues, or to enable search also on inventory services providing metadata without catalogue functionalities. The choice between distributed query and harvesting can be made per data sources. In case of harvesting also the repetition time can be defined per data source. Internally, GI-cat includes several modules (see Figure 13):

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

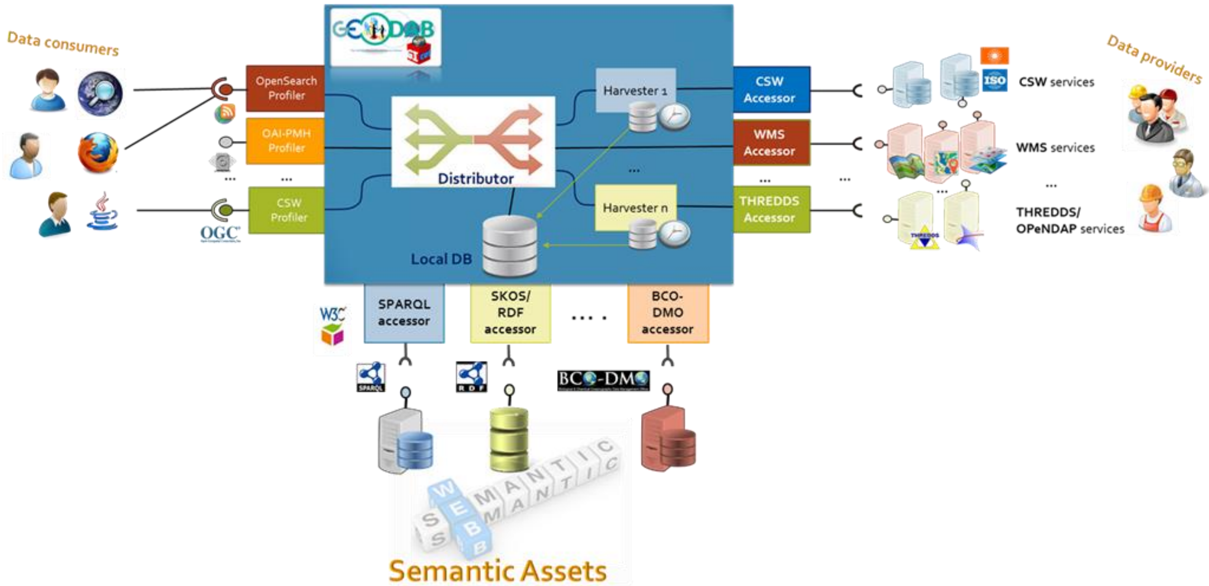

Figure 13 Discovery broker (GI-cat) internal architecture

- *Semantic Enhancement Module (GI-sem)*: a component which implements semantic query expansion [24]. If the semantic query is enabled by configuration, when a query includes a keyword, it is passed as a parameter of a semantic query to a set of connected knowledge bases to search for “related” terms. Each of the resulting term is then used as a keyword in a separate geospatial query. The results are then assembled to provide the complete response to the user. This workflow enables several semantic enhancements depending on the connected knowledge bases, including multilingualism, semantic refinements and suggestions. For example, connecting a multilingual thesaurus supporting English, French, German, Italian, Polish and Spanish, if an user send a request for "moisture" in English, then several separate geospatial queries will be sent through GI-cat, including for "moisture" (English), "humidité" (French), "Feuchtigkeit" (German), "umidità" (Italian), "wilgoć" (Polish) and "humedad" (Spanish). This allows to find datasets annotated in different languages overcoming limitations of syntactic queries on metadata content. GI-sem supports basic relationships such as "related" (i.e. generic relationship; e.g. "soil moisture" is related to "soaking"), "broader" (i.e. generalization; e.g. "soil water" is more general than "soil moisture") or "narrower" (i.e. specification; e.g. "soil moisture" is more specific than "soil water"). GI-sem is implemented through semantic accessors integrated in GI-cat, which map the request to a specific knowledge base interface.

- *Access broker (GI-axe)*: a component which is able to connect with disparate (distributed and heterogeneous) data sources, exposing them through a set of standard catalogue interfaces. By means of data harmonization and protocol adaptation, it is able to download (subset of) datasets from different sources. GI-axe mediates among the connected data sources interfaces, and harmonizes datasets using a small set of internal data models (GI-axe data models). It is also able to carry out on-the-fly transformations for subsetting, reprojection, resampling, encoding. Internally, GI-cat includes several modules (see Figure 14):
    - The *Orchestrator* is in charge of accepting data access requests from the exposed data access interfaces and run the needed workflow for access and transformation. The Orchestrator is a smart component taking into account servers' capabilities: if the original data source already supports the requested transformation, the Orchestrator relies on it, otherwise it calls the Converters.
    - The *Profilers* are adaptors for exposing access interfaces to users. Each Profiler exposes a standard interface carrying out: a) mapping of the data access interface to the internal access interface of the Orchestrator; b) mapping of datasets from the GI-axe data models to the data model of the supported interface, providing also the related encoding. For example the WCS/netCDF Profiler maps the OGC Web Coverage Service (WCS) interface to the internal access interface, and, on the other direction, it transforms the dataset from the GI-axe data model to the netCDF data model and encoding.
    - The *Accessors* are adaptors for connecting data sources. Each Accessor supports a data source carrying out: a) mapping of an internal access request to the interface exposed by the external data source; b) mapping of resulting datasets to the GI-axe data model. For example, the Accessor for FTP hosting GeoTIFF files, maps the data access request to a FTP download request, and on the other direction, it transforms the GeoTIFF dataset to the GI-axe data model.
    - The *Converters* are modules for on-the-fly execution of dataset transformations. These transformations include simple processing aiming not to modify the content of datasets, but only to transform its representation. They include subsetting, reprojection, resampling and encoding. The Converters either use local routines or call external web services exposed through OGC Web Processing Service (WPS) interface.

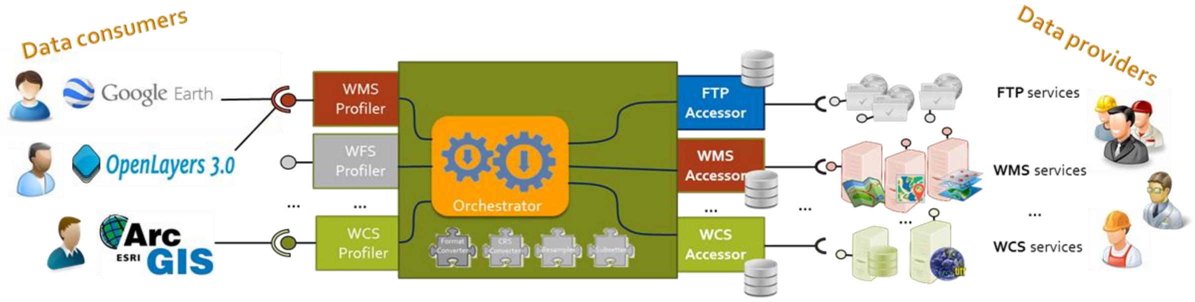

Figure 14 Access Broker (GI-axe) internal architecture

- *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.

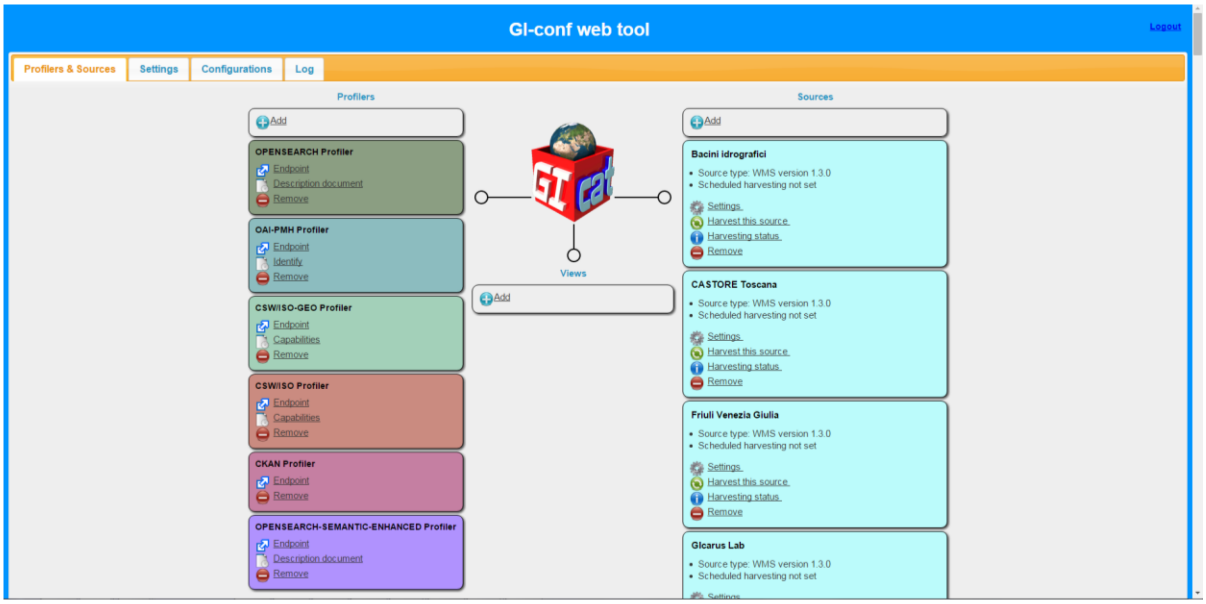

Figure 15 GI-conf screenshot (from the Italian Virtual Hub)

- *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 9 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
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
dee gree Deegree (version 2.2) catalog service	Discovery interface
ESRI ArcGIS Geoportal (version 10) catalog service	Discovery interface
WAF Web Accessible Folders 1.0	Discovery and access interfaces and metadata model
FTP - File Transfer Protocol services populated with supported metadata	Discovery and access interfaces
THREDDS 1.0.1, 1.0.2	Discovery and access interfaces
ncISO THREDDS-NCISO 1.0.1, 1.0.2	Discovery and access interfaces, and metadata model
ncISO THREDDS-NCISO-PLUS 1.0.1, 1.0.2	Discovery and access interfaces, and metadata model
CDI 1.04, 1.3, 1.4 1.6	Discovery interface and metadata model
GI-cat 6.x, 7.x	Discovery and access interfaces
GBIF	Discovery and access interfaces, and metadata model

OpenSearch 1.1 accessor	Discovery interface
OAI-PMH 2.0 (support to ISO19139 and dublin core formats)	Discovery interface and metadata model
NetCDF-CF 1.4	Metadata and data model
NCML-CF	Metadata and data model
NCML-OD	Metadata and data model
ISO19115-2	Metadata model
GeoRSS 2.0	Access interface, and metadata model
GDACS	Access interface, metadata and data models
DIF	Metadata and data model
File system	Access interface
SITAD (Sistema Informativo Territoriale Ambientale Diffuso) accessor	Discovery and access interfaces
INPE	Discovery and access interfaces
HYDRO	Discovery and access interfaces
EGASKRO	Discovery and access interfaces
RASAQM	Discovery and access interfaces
IRIS event	Discovery and access interfaces, metadata model
IRIS station	Discovery and access interfaces, metadata model
UNAVCO	Discovery and access interfaces, metadata model
KISTERS Web - Environment of Canada	Discovery and access interfaces
DCAT	Discovery interface and metadata model

CKAN	Discovery interface and metadata model
HYRAX THREDDS SERVER 1.9	Discovery and access interfaces

Table 9 Preliminary list of Virtual Hub data sources protocols

Table 10 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 10 Preliminary list of protocols supported by Virtual Hub publishing interfaces

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 11), 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 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 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	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	Trees4Future is an Integrative European Research Infrastructure project that aims to integrate, develop and improve major forest genetics and forestry research infrastructures. It will provide the wider European forestry research community with easy and comprehensive access to currently scattered sources of information (including genetic databanks, forest modelling tools and wood technology labs) and expertise.
	Pangaea	The information system PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the operating institutions.
	NSIDC Acadis	The Advanced Cooperative Arctic Data & Information Service (ACADIS) manages data and is the gateway for all relevant Arctic physical, life, and social science data for the National Science Foundation (NSF) Division of

		Polar Programs (PLR) Arctic Research Program (ARC) research community.
	SeaDataNet FP6 project and SeaDataNet2 FP7 project	SeaDataNet objective is to construct a standardized system for managing the large and diverse data sets collected by the oceanographic fleets and the new automatic observation systems. The aim is to network and enhance the currently existing infrastructures, which are the national oceanographic data centres and satellite data centres of European riparian countries, active in data collection. The networking of these professional data centres, in a unique virtual data management system will provide integrated data sets of standardized quality on-line. SeaDataNet CSW interface
	GEOSS (GEO-DAB)	The Group on Earth Observations, GEO, was established by a series of three ministerial-level summits. It currently includes 68 member countries, the European Commission, and 46 participating organizations. The vision of GEO is to create a Global Earth Observation System of Systems (GEOSS) to help realize a future wherein decisions and actions for the benefit of humankind are informed via coordinated, comprehensive and sustained Earth observations and information. The Global Earth Observation System of Systems will provide decision-support tools to a wide variety of users. As with the Internet, GEOSS will be a global and flexible network of content providers allowing decision makers to access an extraordinary range of information at their desk. The IP3 was conceived as a way to exercise the process that has been defined for reaching interoperability arrangements. The 2nd Phase of the AIP will augment the GEOSS Initial Operating Capability previously established.
	GENESI-DR	GENESI-DR, (Ground European Network for Earth Science Interoperations - Digital Repositories), has the challenge of establishing open Earth Science Digital Repository access for European and world-wide science users. GENESI-DR shall operate, validate and optimise the integrated access and use available digital data repositories to demonstrate how Europe can best respond to the emerging global needs relating to the state of the Earth, a demand that is unsatisfied so far.

	GIIDA	GIIDA is a CNR initiative (inter-departmental project) aiming to the design and development a multidisciplinary infrastructure for the management, processing and evaluation of Earth and environmental data. GIIDA aim is to implement the Spatial Information Infrastructure (SII) of CNR for Environmental and Earth Observation data.GIIDA central catalog
	EuroGEOSS	EuroGEOSS demonstrates the added value to the scientific community and society of making existing geographic systems and applications interoperable and used within the GEOSS and INSPIRE frameworks. The project will build an initial operating capacity for a European Environment Earth Observation System in the three strategic areas of Drought, Forestry and biodiversity. The concept of inter-disciplinary interoperability requires research in advanced modelling from multi-scale heterogeneous data sources, expressing models as workflows of geo- processing components reusable by other communities, and ability to use natural language to interface with the models. EuroGEOSS portal
	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 11 List of infrastructures and initiatives using the GI-suite Brokering Framework

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

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

6.2.2 Integration of the GI-suite Brokering Framework with the SRP GIS-Broker

The GIS-Broker is a software tool developed by SRP for enabling advanced Geographical Information Systems (GIS) applications. It supports public and private (open) data providers to overcome the high entry barriers to data sharing, allowing to easily implement standard-compliant Spatial Data Infrastructures (SDIs) on top of map and feature data sources. The GIS-Broker provides functionalities for OGC/INSPIRE-compliant metadata administration, GIS-analysis, processing, transformation and production, map and feature management, user and resources profile definition, etc. It supports the connection with different data sources, including OGC WMS, OGC WFS and legacy databases, through dedicated adaptors.

While the GI-suite Brokering Framework focuses on mediation between heterogeneous data sources from different communities (Earth Observation, scientific disciplinary Communities-of-Practice, observation networks, crowdsourcing systems, etc.), the GIS-Broker focuses on the support of the entire information life-cycle for a well-defined class of stakeholders which typically are public authorities providing geographical information for public services and companies.

The integration of the GIS-Broker with the GI-suite Brokering Framework allows to assure advanced functionalities to data providers and, at the same time, facilitate the use of published open data and their integration by a wider class of users.

Since both the technological solutions provide server-side APIs, a loosely-coupled integration is easily implemented. In particular, the GIS Broker exposes standard interfaces for discovery (OGC CSW), access (OGC WFS) and visualization (OGC WMS) which are already supported by the GI-suite Brokering Framework. This makes possible for the GI-suite Brokering Framework to expose data published by the GIS-Broker as one of the connected data sources.

Figure 16 shows the component diagram of the integration between the GI-suite Brokering Framework and the GIS-Broker. On top, the figure shows an example of Local SDI including a set of possibly heterogeneous data sources and a management GUI. In the Virtual Hub, the GIS-Broker is able to connect all the data sources providing management services accessible through the management GUI. It exposes OGC compliant interfaces for discovery, feature access and map visualization. The GI-suite Brokering Framework connects with the GIS-Broker and with other data sources (only GEOSS, INSPIRE and Copernicus are shown for the sake of clarity) exposing its capabilities through simple GI APIs for app development. Comparing Figure 16 and Figure 9, it should be noted that: a) the VH package in Figure 16 collects the components hosted by the VH node of Figure 9 (for the sake of clarity, Figure 16 only shows the relevant components); b) the GI-suite Brokering Framework of Figure 16 implements the Brokering Framework package of Figure 9; c) the GIS-Broker of Figure 16 implements the SDI Manager of Figure 9.

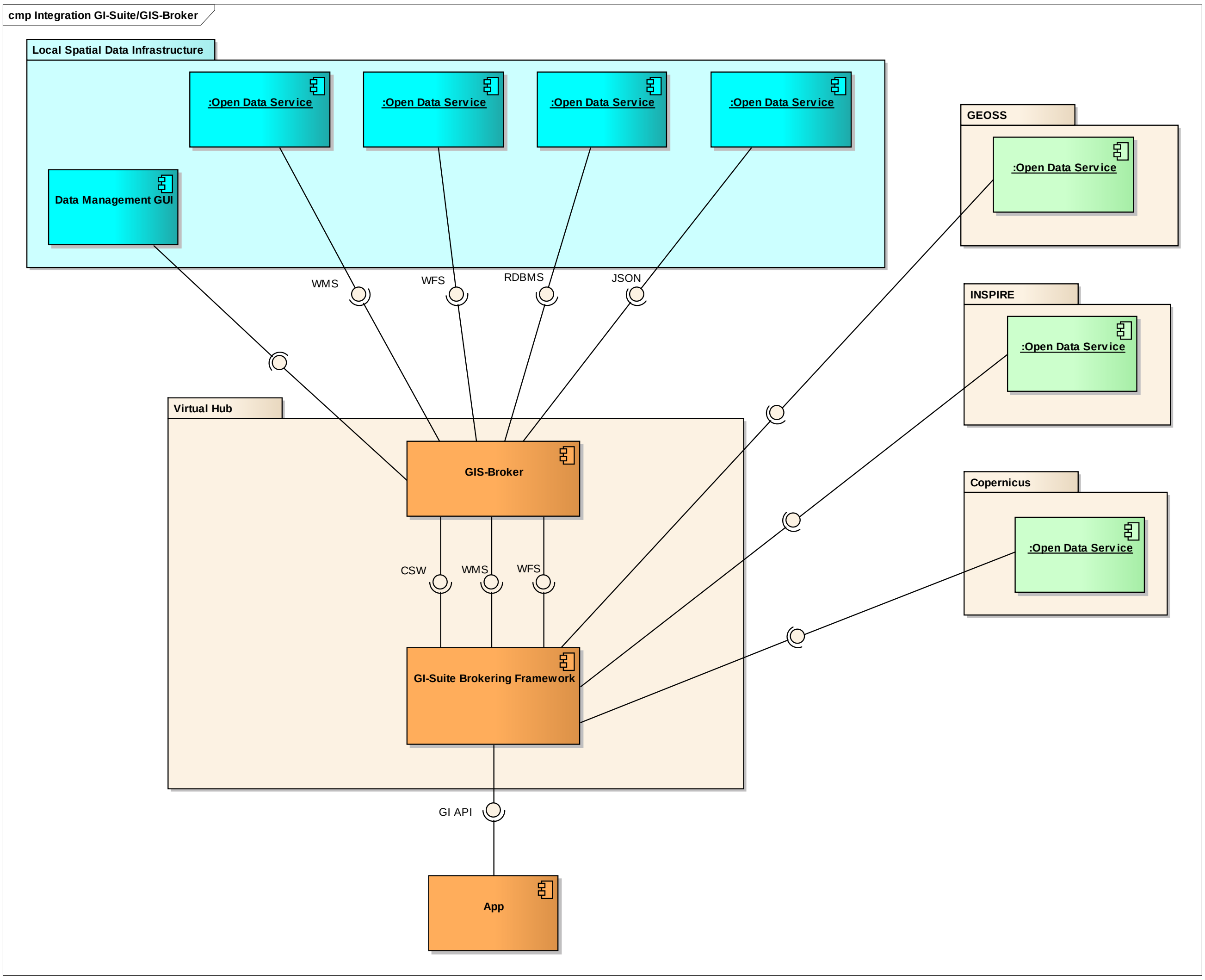

Figure 16 Integration of the GI-suite Brokering Framework and GIS-Broker

6.2.3 Integration of the GI-suite Brokering Framework with the BRGM SOS Publishing Service

The ENERGIC OD application for Coastline Evolution Monitoring developed by BRGM includes a specific use-case dedicated to “create observations to be validated” through a mobile application [29]. Therefore, it requires support for uploading observations by users. The solution chosen by BRGM for implementing such functionality is the OGC Sensor Observation Service standard and technology, where the mobile application is a captor (see Figure 17). The mobile client application will integrate the Sensiasoftware[56] library implementing the needed methods (drivers) to support two different kinds of acquisitions:

a) Writing a coastline by walking the see border,
b) Giving an alert with picture and data about a coastline change.

On the server side, a SOS service will be implemented using the OpenSensorHub[57] technology. After a validation phase by a Coastline specialist, if the uploaded dataset is approved, it will be accessible as a public open data through OGC WFS interface. Thanks to the GI-suite Brokering Framework it will be published by the Virtual Hub.

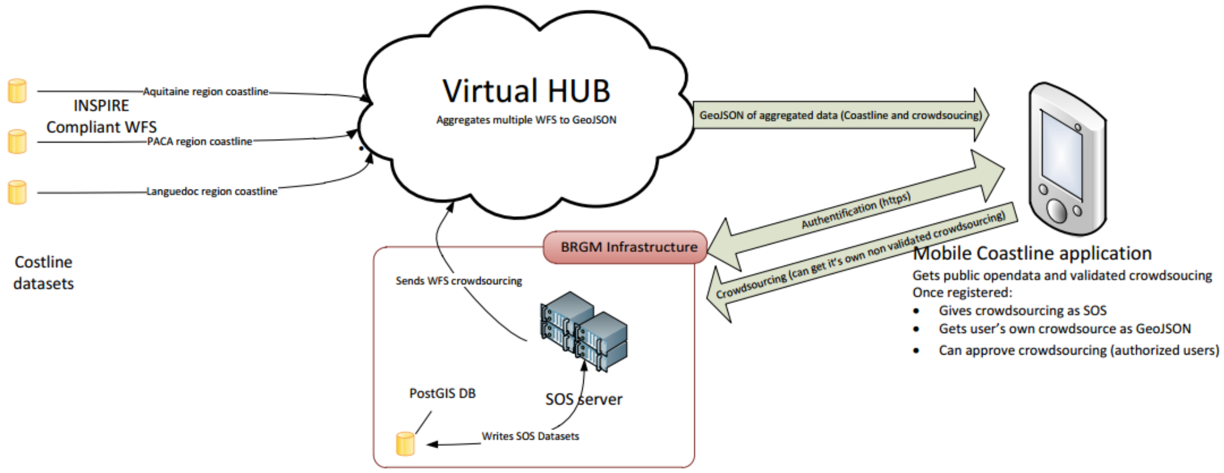

Figure 17 Crowdsourcing support in the BRGM Coastline Evolution Monitoring application

According to the architecture of the BRGM Coastline Evolution Monitoring application, Figure 17 shows the SOS Server as an autonomous system managed by BRGM and accessed by the Virtual Hub. However, such an application can be considered more generally as a proof of concept for scientific open data publication and crowdsourcing. Many other potential applications may require data publishing without having existing autonomous systems available for upload. Therefore, the solution developed by BRGM is a potential upgrade of the Virtual Hub itself. The SOS Server could become part of the Virtual Hub architecture playing the role of the Data Publisher and the Local Storage of Figure 10.

7 DEPLOYMENT

7.1 Deployment plan

The ENERGIC OD DoW [30] states that the project will realize five national VH in France, Germany, Italy, Poland and Spain. Moreover, a sixth VH at European level will be optionally implemented if considered useful, possibly to aggregate the five national VH (Figure 18).

During an internal workshop held in Florence [31], the project partners discussed the option that, by a

[56] http://www.sensiasoftware.com/
[57] http://docs.opensensorhub.org

marketing point-of-view, it might 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. However, as discussed in later consortium meetings, the original idea of national-level Virtual Hubs has strong elements in support. In particular, some data sources, namely INSPIRE, are already structured at national level. Moreover, having national level VHs may help to overcome linguistic issues.

Although the final decision was to keep the national-level structure that the project was committed to provide, this discussion stressed the importance of keeping the architecture as loose as required to support many different topologies (see architectural principles in section §4.5).

During the first year of activities, the partner SRP proposed to have a further deployment of a VH serving the Berlin metropolitan area, possibly connected to the German VH. Since this deployment is compliant with the commitment to provide national-level VHs, but also give the possibility to test VH chaining, it was agreed to have a sixth Virtual Hub managed by SRP.

The deployment plan is then shown in Figure 18.

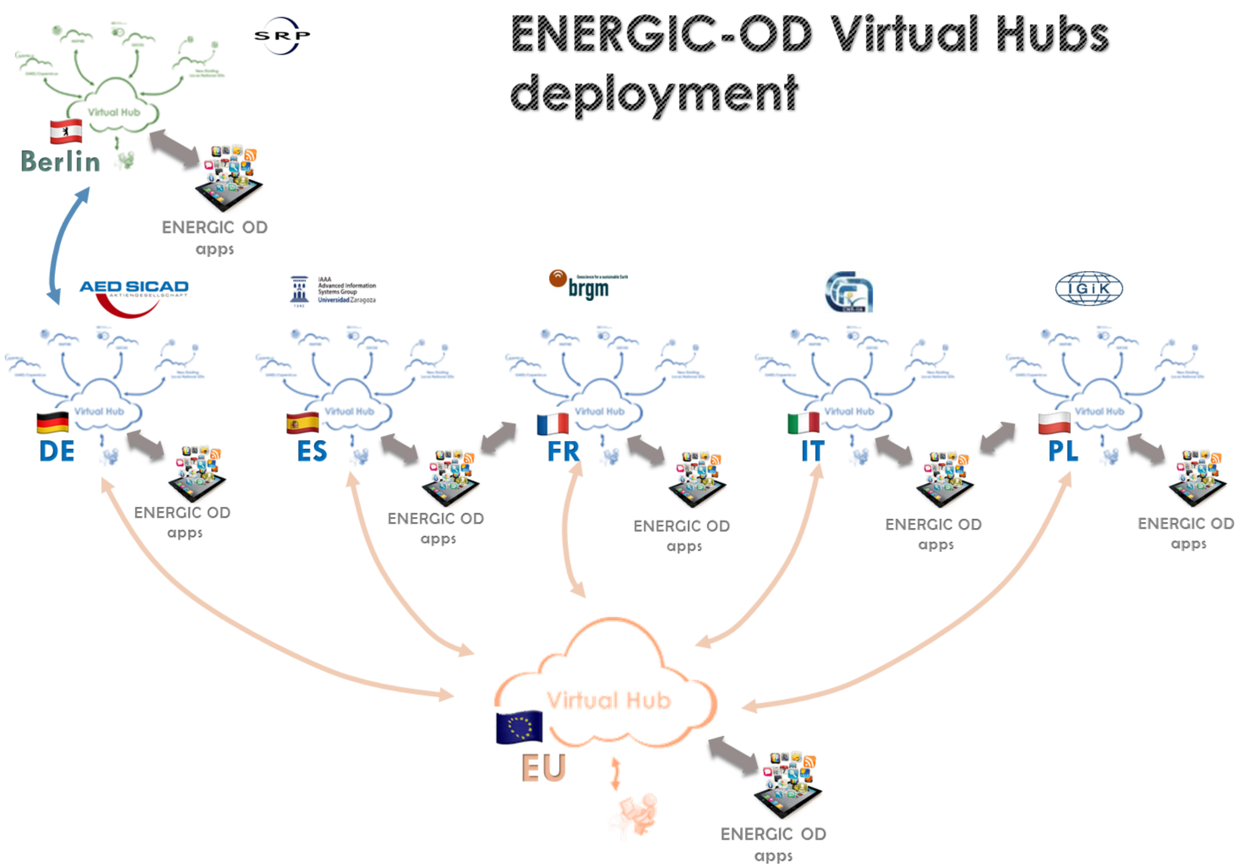

Figure 18 Deployment plan and status (at month 16)

7.2 Deployment status

The deployment phase officially started on month 13. However, most of the VH managers anticipated the activities to familiarize with technologies and in preparation to app development. The first action of each VH manager was a decision on the adopted deployment strategy (Table 12 and Figure 19).

Virtual Hub	VH Manager	Choice
VH Berlin	SRP	Local Installation
VH France	BRGM	Local Installation
VH Germany	AED-SICAD	Public cloud (Amazon)
VH Italy	CNR-IIA	Public cloud (Amazon)
VH Poland	IGIK	Virtualized local installation
VH Spain	UNIZAR	Virtualized local installation

Table 12 Deployment choices

Based on these decisions, the software package for the first iteration has been provided to the Consortium in three different formats:

- As a Java Web application ARchive (WAR) for deployment on Servlet containers for local installation
- As an AMI (Amazon Machine Images) that contains the information required to launch an instance, which is a virtual server in the Amazon cloud;
- As a virtualised image in Open Virtualization Format (OVF) which is an open standard for packaging and distributing virtual appliances or, more generally, software to be run in virtual machines.

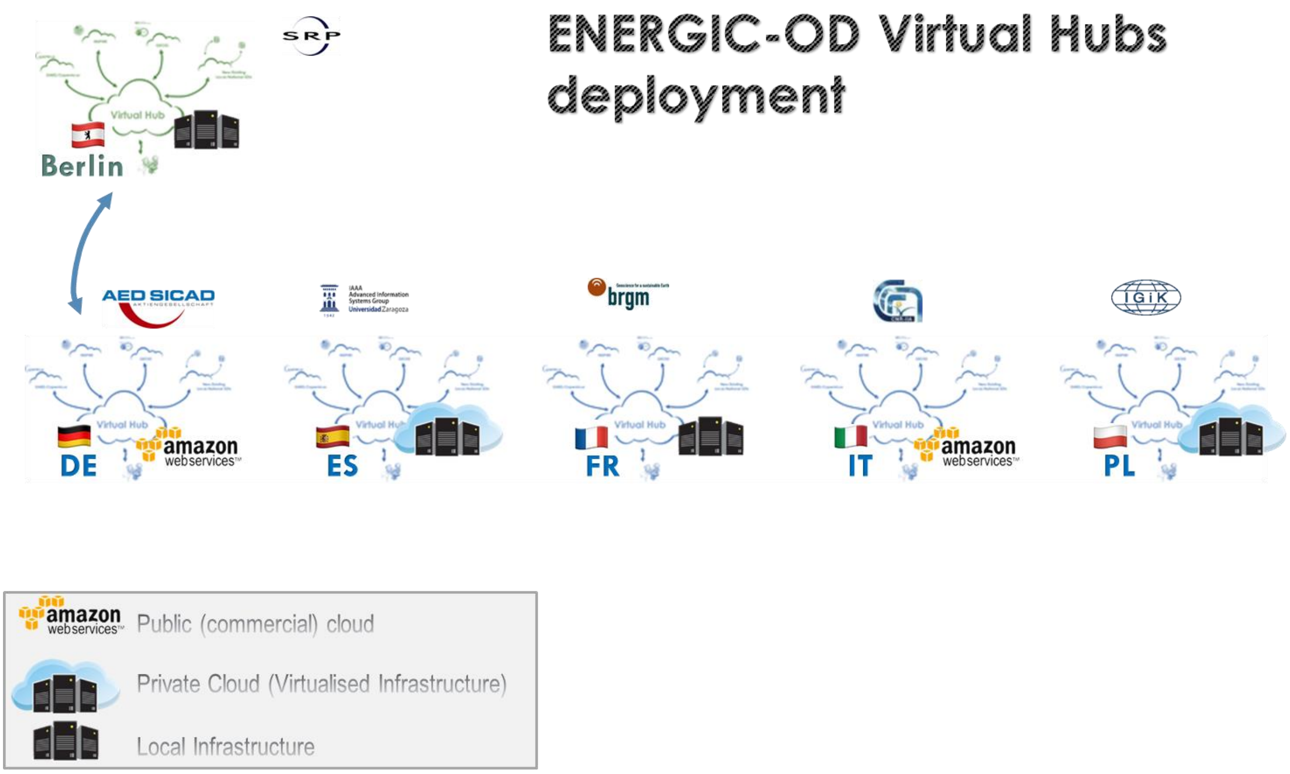

Figure 19 Deployment strategy for local and national Virtual Hubs

7.3 Virtual Hub data sources

The local and national Virtual Hubs provide access to the set of geospatial open data sources supporting the development of potential applications by their users and primarily by the ENERGIC-OD app developers. The selection of data sources has been carried out by the WP3 (Open Data Survey) and WP6 (Development of new innovative applications). Table 13 Preliminary data sources to be connected at each Virtual Hub shows a list of platforms and portals identified by WP3 and WP6 activities. They are considered as a preliminary list of potential data sources to be connected to the local and national Virtual Hubs.

Italy	Il portale del Governo Italiano sui dati aperti	http://www.dati.gov.it/	National Level
Italy	Portale Cartografico Nazionale (National Cartographic Portal)	http://www.pcn.minambiente.it/GN/	National Level
Italy	Open data Italy	http://www.datiopen.it/	National Level
Italy	Open data Trentino	http://dati.trentino.it/	Local Level
Italy	Portale del Servizio Geologico d'Italia	http://www.isprambiente.gov.it/it	National Level
Poland	geoportal Poland	http://www.geoportal.gov.pl/start	National Level
Poland	Dane Publiczne	https://danepubliczne.gov.pl/pl/	National Level
Germany	Transparenzportal Hamburg	http://transparenz.hamburg.de/	Regional Level
Germany	Datenportal für Deutschland	https://www.govdata.de/	National Level
Germany	Open data portal Rostock	http://www.opendata-hro.de/	Regional Level
Germany	Open data catalogue Germany	https://offenedaten.de/dataset	National Level

Germany	Open data Bayern	https://opendata.bayern.de/?3	Regional Level
Germany	Berlin Open Data	http://daten.berlin.de/datensaetze	Regional Level
Germany	Transparenzportal Bremen	http://transparenz.bremen.de/	Regional Level
Germany	Geoportal Germany	http://www.geoportal.de/DE/Geoportal/geoportal.html?lang=de	National Level
Spain	Portal open Data DA	http://abertos.xunta.es/portada	Regional Level
Spain	DIE-E	http://www.idee.es/	National Level
Spain	Datos Abiertos de Zaragoza	http://www.zaragoza.es/ciudad/risp/	Local Level
Spain	OpenGov	http://opengov.es/	National Level
Spain	Gencat Open Data	http://dadesobertes.gencat.cat/en	Regional Level
Spain	Aragón Open Data	http://opendata.aragon.es/	Regional Level
Spain	Open Data Euskadi	http://opendata.euskadi.eus/w79-home/eu/	Regional Level
Spain	Datos.gob.es	**http://datos.gob.es/**catalogo	National Level
France	data gouv France	https://www.data.gouv.fr/fr/	National Level
France	Nantes Open Data	http://data.nantes.fr/	Regional Level
France	Geo Bretagne	http://cms.geobretagne.fr/opendata	Regional Level
France	Data Loire Atlantique	http://data.loire-atlantique.fr/	Regional Level
France	Data Pays de La Loire	http://data.paysdelaloire.fr/	Regional Level
France	Rennes Metropole	http://www.data.rennes-metropole.fr/	Regional Level

Table 13 Preliminary data sources to be connected at each Virtual Hub

In some cases further exploration has been carried out beyond the results documented in WP3 and WP6 deliverables. In particular:

- For Spain, UNIZAR identified more than one hundred open data platforms and SDIs at both local and regional levels. VH-ES managers will be initially connecting to the data sources (platforms) included in the table. After that, they will check if the catalogues of regional-level platforms are integrated into the national OD platform or national SDI, to connect the not-integrated ones directly to the VH. A preliminary assessment of the many local level platforms identified showed that many of them are not properly maintained. In its analysis, UNIZAR assumed that the ones maintained and correctly managed are integrated into the catalogues of their respective regional platforms.

- For Italy, POLIMI identified several hundreds of open datasets published by public authorities at different administrative level in Italy (municipalities, provinces, regions). Further work is needed to investigate whether they are available through the catalogue service of existing open data platforms. In case of affirmative answer the identified open data platform will be connected to the Italian VH, otherwise the datasets will be individually published through the Italian VH, making them discoverable by the open data community.

7.4 Virtual Hub naming

The ENERGIC OD Consortium decided to define a uniform naming schema for the Virtual Hubs, under the official project web domain (energic-od.eu). The schema naming is as follows:

`vh-`*`xyz`*`.energic-od.eu`

where *`xyz`* is equal to `berlin` for the local-level VH in Berlin and to the country code for national-level VHs. The decision to use country codes depends on the need to have a global naming, and at the same time avoiding to use a unique language (e.g. English) for the national-level VHs that mainly address national users.

The resulting naming of VHs is then:

Virtual Hub	Name
Berlin Local Virtual Hub	`vh-berlin.energic-od.eu`
French National Virtual Hub	`vh-fr.energic-od.eu`
German National Virtual Hub	`vh-de.energic-od.eu`
Italian National Virtual Hub	`vh-it.energic-od.eu`
Polish National Virtual Hub	`vh-pl.energic-od.eu`
Spanish National Virtual Hub	`vh-es.energic-od.eu`

8 REFERENCES

[1] IEEE, SWEBOK v3 - Guide to the Software Engineering Body of Knowledge, P. Bourque e R. E. Fairley, A cura di, IEEE, 2014.

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

[3] 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.

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

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

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

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

[8] J. Ernst, «Methodology and a list of criteria for the assessment of open data platforms and initiatives across the EU - Deliverable D3.1,» ENERGIC-OD Consortium, 2015.

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

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

[11] ENERGIC OD Consortium, «Annex I - Description of Work (Amendment 2015),» 2015.

[12] S. Braumann e H. Kuechly, «Application based requirements and standards catalogue - Deliverable D6.1 (second release),» ENERGIC OD Consortium, 2016.

[13] R. Brumana, «Report on the most relevant fields of application and user needs, strictly related to innovative applications, services and semantic content to be implemented - Deliverable D4.2,» ENERGIC OD Consortium, 2015.

[14] R. Brumana, «Report on Data Registration, Integration and Harmonization Methodology aligned with the INSPIRE directive and on-going initiatives - Deliverable D4.1,» ENERGIC OD Consortium, 2016.

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

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

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

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

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

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

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

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

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

[24] 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.

[25] M. A. Latre, «R&D State-of-the-art report - Deliverable D2.3,» ENERGIC OD Consortium, 2015.

[26] P. Colpaert, A. Dimou, G. A. Atemezing e R. Troncy, «Technical toolset for open data - Deliverable D4.5,» Apps4EU Consortium, 2014.

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

[28] 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/.

[29] S. Braumann e H. Kuechly, «Implementation Plan - Deliverable D6.4,» ENERGIC OD Consortium, 2016.

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

[31] P. Mazzetti, «Report on WP5 Workshop - Deliverable D5.9,» ENERGIC OD Consortium, 2015.

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

ANNEX I – SUMMARY OF CHANGES AND UPDATES

The present annex provides a summary and justification of the main changes between version 1 and version 2 of the deliverable D5.1 on "ENERGIC OD Virtual Hub – System Architecture".

- The identification of VH actors in section 3.1 (Actors) has been approved by WP8 lead organization (Trilateral) for harmonization with exploitation activities.
- Section 2.3 (Open Data in ENERGIC OD) has been revised by WP3 lead organization (GeoKomm) for harmonization with work in WP3 (Open Data Survey).
- Section 3.2 (User requirements) has been revised by WP4 and WP6 lead organizations (POLIMI and LUP) for harmonization and inclusion of outcomes from WP4 (Requirements and specifications: SDI, data harmonisation and applications addressing user needs) and WP6 (Development of new innovative applications) activities in the first year of ENERGIC OD.
- Section 3.4 (System Requirements) has been revised by WP4 and WP6 lead organizations (POLIMI and LUP) for harmonization and inclusion of outcomes from WP4 (Requirements and specifications: SDI, data harmonisation and applications addressing user needs) and WP6 (Development of new innovative applications) activities in the first year of ENERGIC OD. Table 3 (ENERGIC OD system requirements) has been updated correspondingly.
- Section 5.2 (Enterprise Viewpoint) has been revised by the WP6 lead organization (LUP) to include outcomes related to app supporting scenarios.
- Section 5.6 (Technology Viewpoint) has been revised by WP2 lead organization to include outcomes of WP2 (State of the art in R&D, projects & technologies).
- Section 6.2 (System integration) has been revised including a description of the first components included in the VH architecture and contributions from the first activities on integration of components in ENERGIC OD.
- Section 7 (Deployment) has been updated with plans and status at month 16 of the project.