VIRTUAL HUBS: A BROKERED ARCHITECTURE FOR FACILITATING OPEN DATA SHARING AND USE

ROBERTO RONCELLA - ROBERTO.RONCELLA@CNR.IT

NATIONAL RESEARCH COUNCIL OF ITALY – INSTITUTE OF ATMOSPERIC POLLUTION (CNR-IIA)

OUTLINE

1. ## Theoretical Section
   - Introduction/Main concepts
   - Geospatial Data Challenges
   - Virtual Hub
   - Pan-European Virtual Hub

2. ## Pratical Section
   - Demonstration of Virtual Hub Test Portal
   - Demonstration of Pan-European Virtual Hub

INTRODUCTION

GEOSPATIAL INFORMATION 1/2

- It is a key resource for many applications

- It is growing in size and variety at an exceptionally fast rate

- New Satellite and remote-sensing system give the opportunity to enhance our knowledge of the Earth System

- Open Data movement is making geospatial data available

Wonderful…but

GEOSPATIAL INFORMATION 2/2

It also poses great challenges to scientist and information technology experts

- Geospatial Open Data (GEO OD) are not easily usable by developers who are not expert in geospatial science and technologies

- The society is not able to fully exploit the potential of GEO Open Data

EARTH OBSERVATION (EO) DATA

Two main concepts:

○ **Open Data**

○ **Big Data**

10/12/2018   roberto.roncella@cnr.it

OPEN DATA

«Idea that certain data should be freely available to everyone to use and republish as they wish withouth restrictions from copyright, patents or other mechanisms of control»

The key features of openness are:

- Availability and access: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form

- Reuse and redistribution: the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable

- Universal participation: everyone must be able to use, reuse and redistribute — there should be no discrimination against persons or groups

BIG DATA ('V' axes)

- **Volume:** big datasets or large amount of datasets

- **Variety:** great heterogeneity

- **Velocity:** efficient dataset handling

- **Veracity:** the need of documenting quality and uncertainty

- **Visualization:** the need of presenting complex data structure in an efficient way

GEOSPATIAL DATA CHALLENGES

MAIN BARRIERS TO THE USAGE OF GEOSPATIAL OPEN DATA (GEO-OD)

1. **Governance:** geospatial OD systems have different mandate and will evolve autonomously
   - Multiple data providers
   - Heterogeneous policies

2. **Interoperability:** geo OD are highly heterogeneous
   - Data/metadata model and formats
   - Coordinate Reference Systems
   - Data services specifications
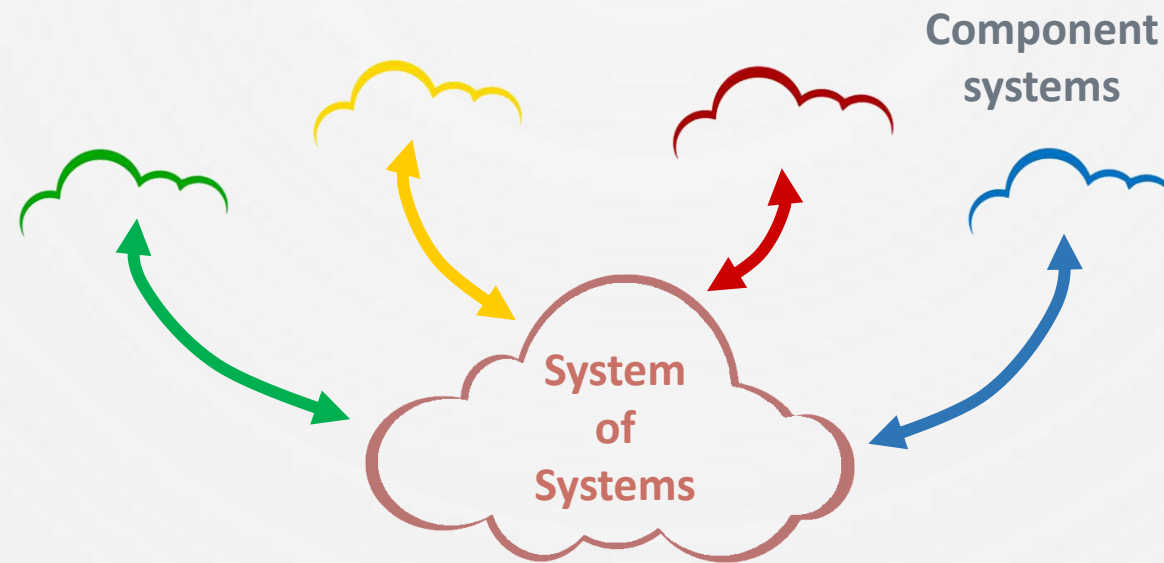   - Quality and reliability
   - Semantics

3. **Usability**
   - Complex service interfaces
   - Complex metadata/data models/encodings
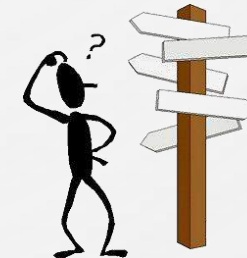
1. ADDRESSING GOVERNANCE: SYSTEM OF SYSTEMS (SoS)

Component systems

System of Systems

SYSTEM OF SYSTEMS (SoS): POSSIBLE APPROACHES

o **Federated Approach**

o **Brokered Approach**

FEDERATED APPROACH

All the local systems agree on a common model (e.g. standards and/or tools)

- Need to define the common model (complex in multidisciplinary environments)
- Need to impose the adoption of the common model
- Need to implemet the common model (high IT expertise)

Can it works?

- Controlled environments (with a central authority)
- Systems with strong interest to partecipate in the federation

BROKERING APPROACH 1/2

Specific third-party components (brokers) mediate between heterogeneous

systems

- Need to develop brokers (complex)

- Need to deploy brokers

- Need to avoid that brokers become single points-of-failure/bottlenecks

- Need to define brokers governance

Can it works?

- Uncontrolled environments (e.g. the web)

- Systems where systems do not have strong interest to participate in the SoS

BROKERING APPROACH 2/2

Brokering it is not against standardization

- Standardization is useful to reduce heterogeneity as much as possible (do not reinvent the wheel, leverage previous efforts by experts in data and service modelling), especially in disciplinary communities.

- Brokering resolve the remaining and irreducible heterogeneity, especially in multidisciplinary contexts.

2. ADDRESSING INTEROPERABILITY: BROKERED SYSTEM OF SYSTEMS (SoS)

Component systems

brokers
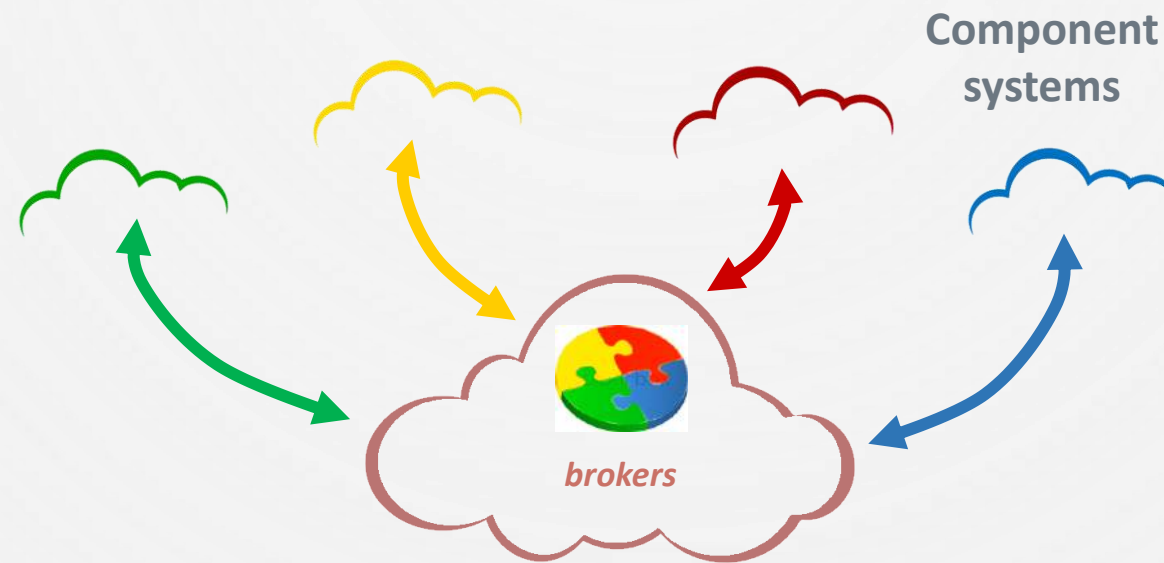
3.   ADDRESSING USABILITY: PORTAL AND APIs
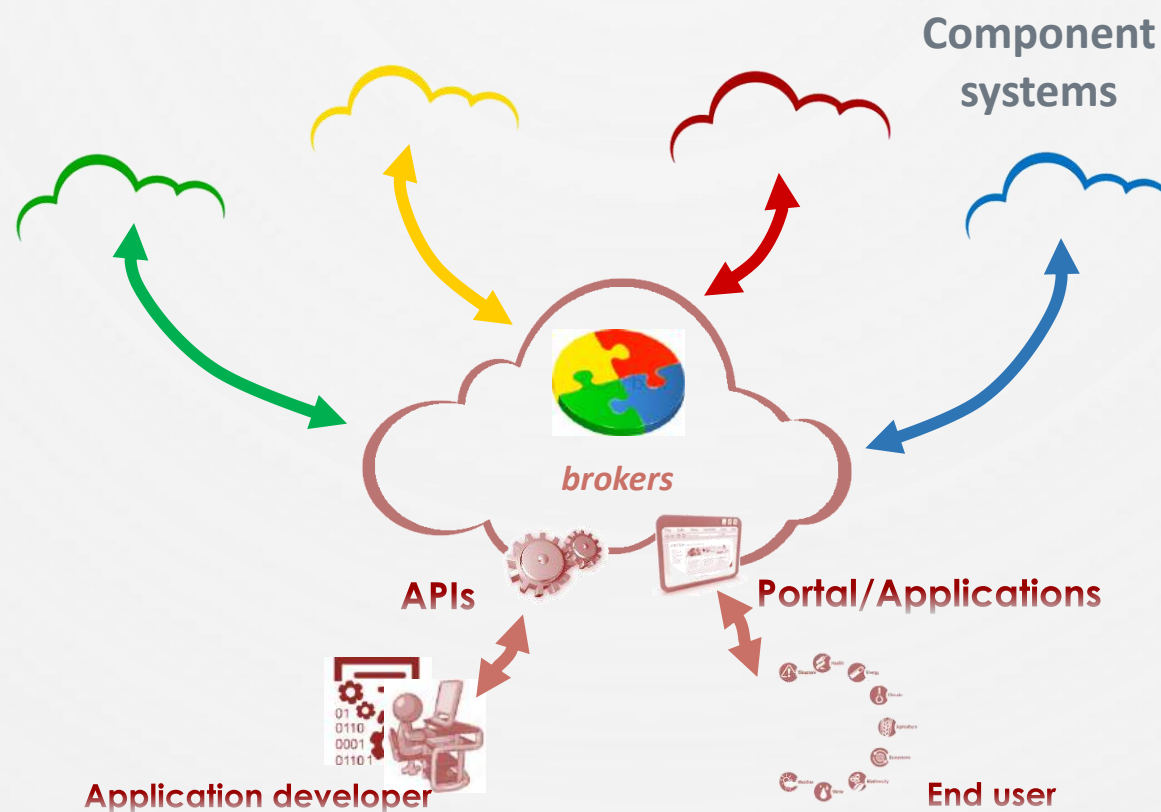
Component systems

brokers

APIs

Portal/Applications

Application developer

End user

VIRTUAL HUB

WHAT IS A «VIRTUAL HUB»?

- *hub-and-spoke* distribution paradigm:

  ○ "*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**"*

  ○ In software engineering *hub-and-spoke* architectures are based on the *(message) broker* pattern

VIRTUAL HUB CONCEPT

Component systems

Virtual Hub

Upstream services
Resource provision

Midstream services
Interoperability, Policy harmonization

Downstream services
User services

- Brokering Framework

Virtual Hub APIs

Virtual Hub Portal/Applications

- Standard interfaces
- Restful API
- Web API

Application developer

End user
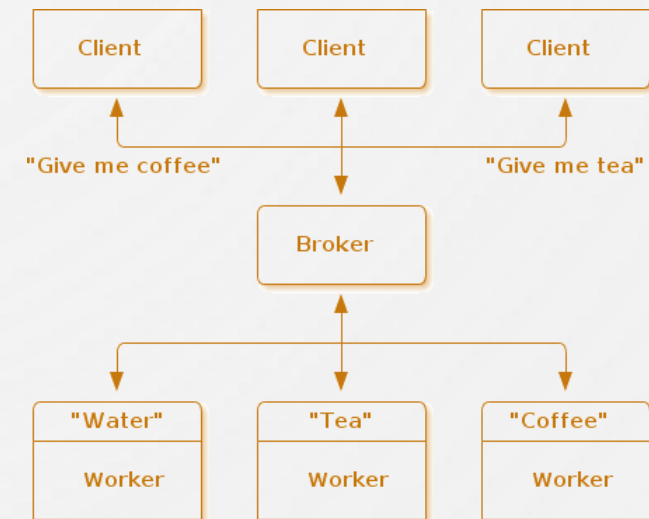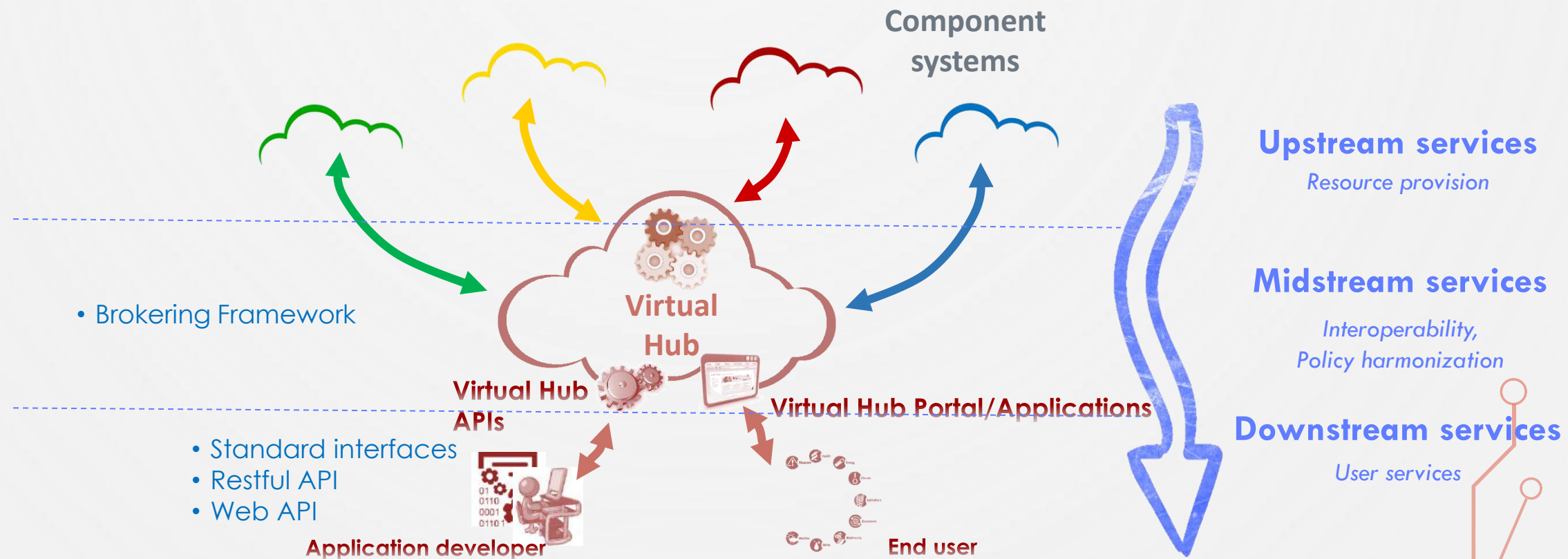
VIRTUAL HUB - APIs

More powerful

- **Geospatial standard interfaces**: well-known standard interfaces. E.g. OGC CSW, OGC WxS, OpenSearch, etc.

- **RESTful APIs**: supporting machine-to-machine interaction with the VH through the exchange of JSON messages through HTTP requests and responses

- **Web APIs**: Javascript library providing easy access to the most common functionalities, and integration with other widespread libraries (e.g. openlayers)

More user-friendly

10/12/2018   roberto.roncella@cnr.it
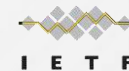
VIRTUAL HUB FUNCTIONALITIES

○ Harmonized data discovery

    ○ Discovery interface mediation

    ○ Metadata harmonization

○ Harmonized data access

    ○ Access interface mediation

    ○ Data harmonization (reprojection, subsetting, interpolation, format encoding)

○ Advanced services

    ○ Semantic queries, metadata multilingualism

SEMANTIC QUERIES

Semantic enhancements to discovery through geospatial query expansion based on external knowledge bases

Three main relationships:

- **Broader:** assert that one concept is broader in meaning (i.e. more general) than another

- **Narrower:** assert that one concept is narrower in meaning (i.e. more specific) than another

- **Related:** assert an associative relationship between two concepts

10/12/2018   roberto.roncella@cnr.it

VIRTUAL HUB TEST INSTANCE: DATA SOURCES TYPE

- What kind of data/services are retrieved?

  ✓ Satellite Data (Sentinel Data/ Landsat Data)
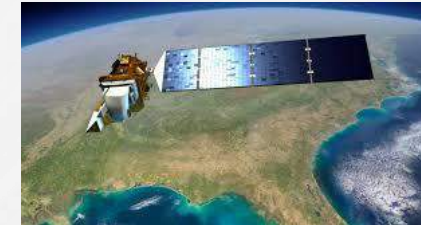
  ✓ Shape files

  ✓ **OGC Web Map Service** (e.g. Gicarus Lab)
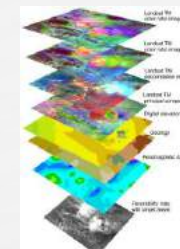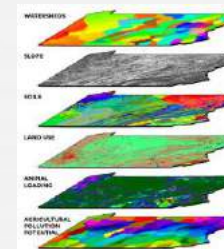
OGC WEB MAP SERVICE (WMS)

- **Maps of spatially referenced data dynamically** from geographic information.

- A "map" is a portrayal of geographic information as a digital image file suitable for display on a computer screen (e.g. png, gif, jpg, …)

- A basic WMS classifies its geographic information holdings into **"Layers"** and offers a finite number of predefined **"Styles"** in which to display those layers.

Layers
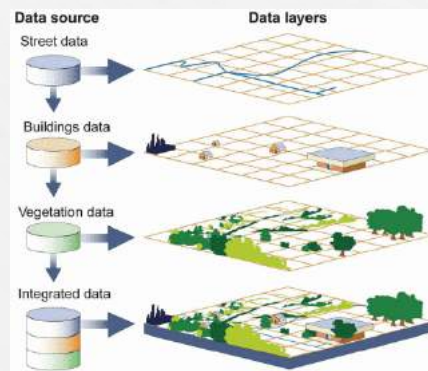
Styles
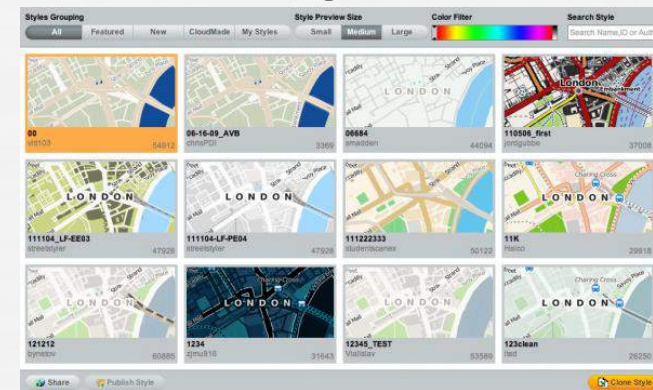
WMS SERVICE MAIN OPERATIONS

- GetCapabilities (Mandatory)
  - Available layers
  - Available styles
  - Provider info

- GetMap (Mandatory)
  - Retrieve portion of the map according to specified parameters:
    - Layer
    - Style
    - Width & height
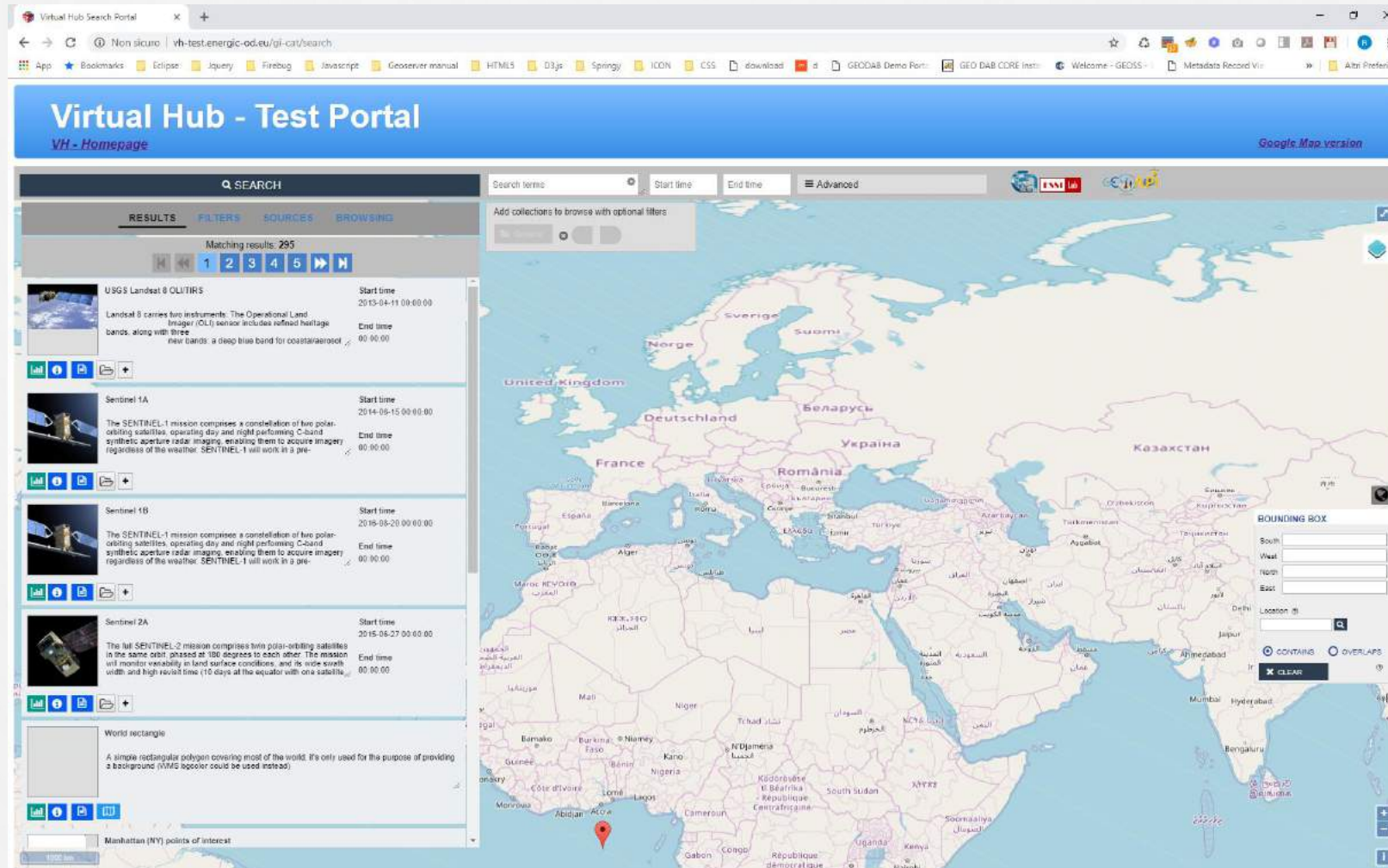   - Bounding box
    - Image format

WMS CLIENTS
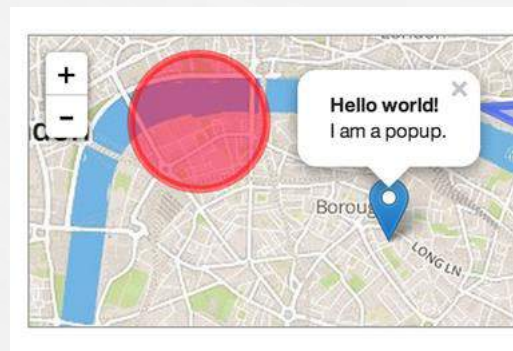
VIRTUAL HUB TEST INSTANCE

THE ENERGIC-OD PAN-EUROPEAN VIRTUAL HUB

EUROPEAN NETWORK FOR REDISTRIBUTING GEOSPATIAL INFORMATION TO USER COMMUNITIES – OPEN DATA (ENERGIC OD)
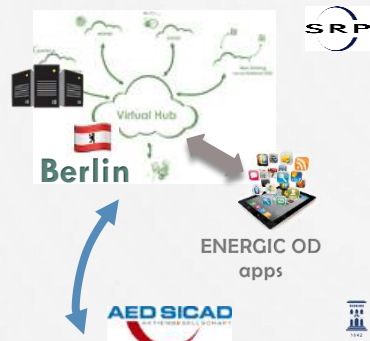
➢ Objective:

- Development of virtual hubs that facilitate the use of open (freely available) geographic data from different sources for the creation of innovative applications and services

➢ Specific activies

- Development of Virtual Hub technology
- Development of (10) pilot applications

ENERGIC OD VIRTUAL HUB DEPLOYMENT

Regional VH

National VHs

pan-European VH

THE ENERGIC OD VIRTUAL HUB

- Virtual Hub distribution:
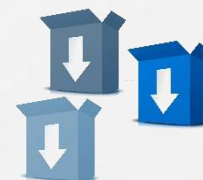
    - **Running instances** (for developers and end-users)
        - 1 regional VH
        - 5 national VHs
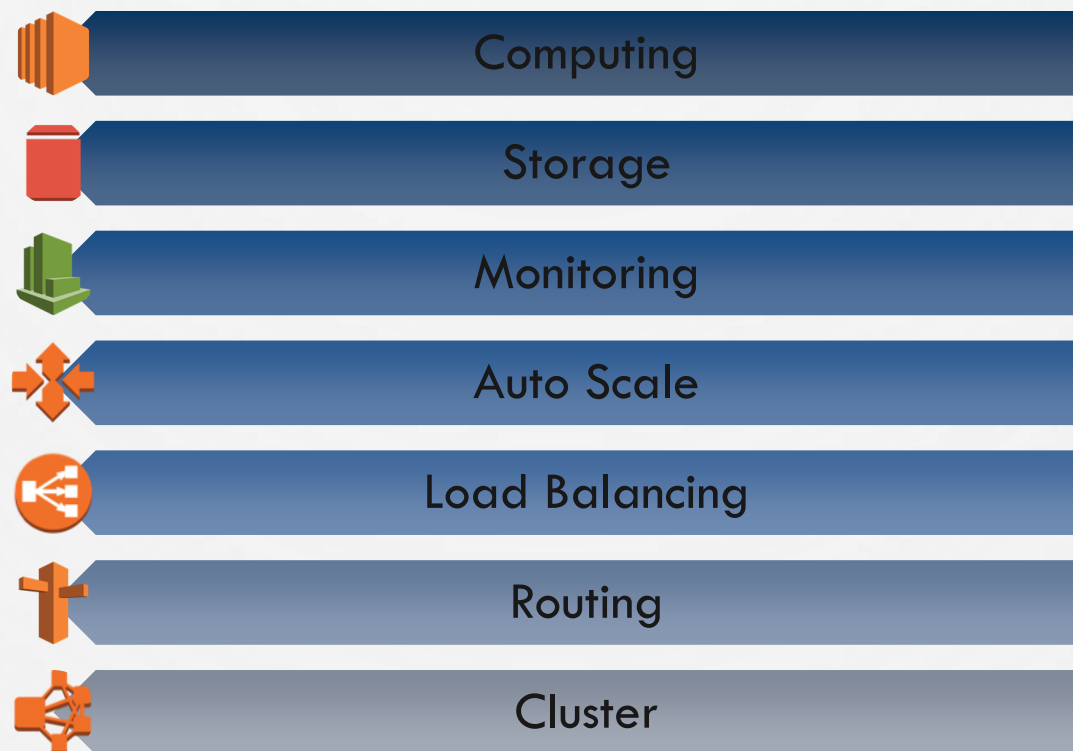   - 1 pan-European VH

    - **Software packages** (for service providers and system integrators)
        - Web archives for servlet containers
        - Virtual machines for deployment on public/private clouds

VIRTUAL HUB CLOUD SERVICES

	Computing
	Storage
	Monitoring
	Auto Scale
	Load Balancing
	Routing
	Cluster

VIRTUAL HUB IN CLOUD SERVICE MODEL
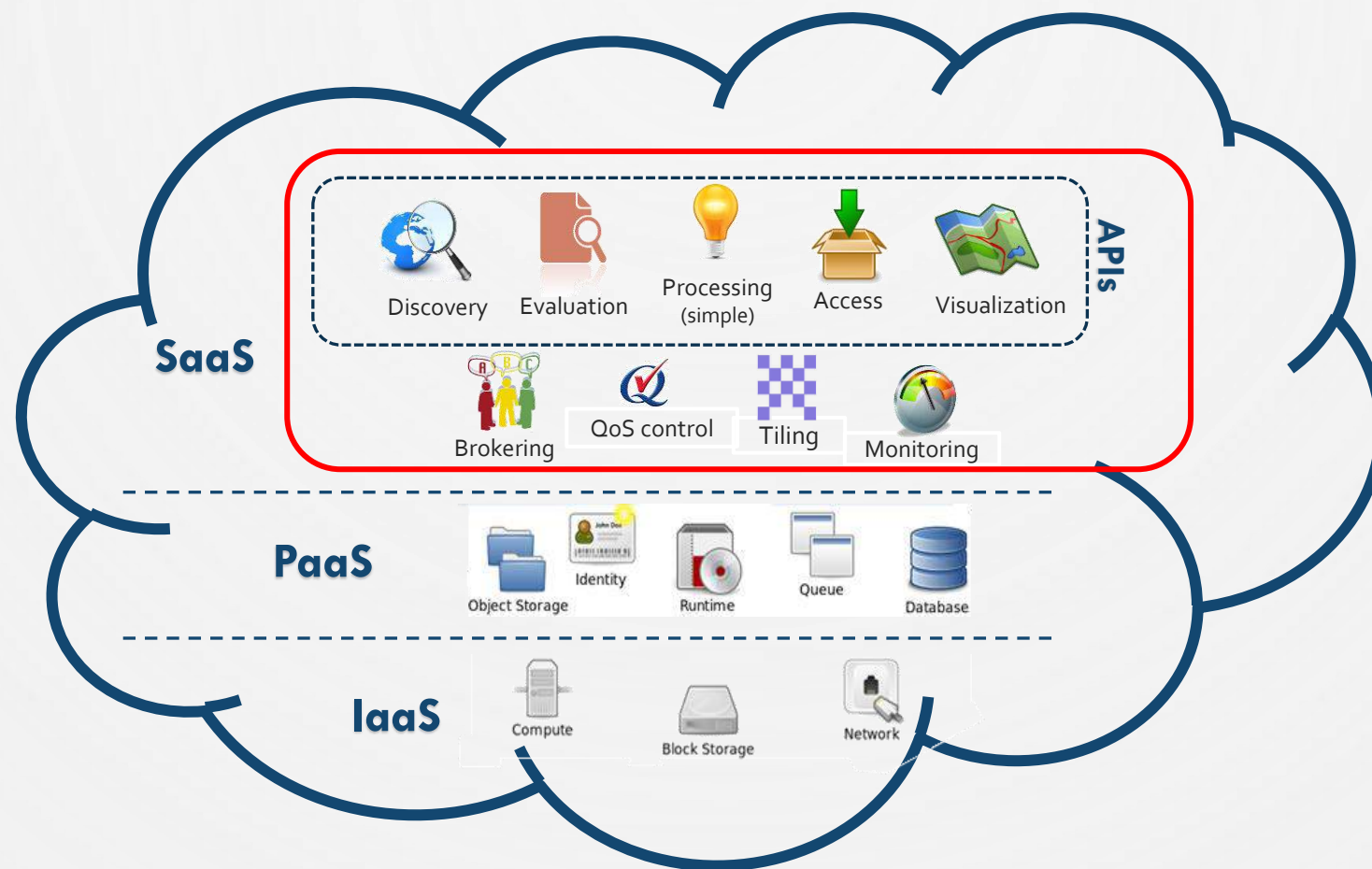
THE PAN-EUROPEAN VIRTUAL HUB PORTAL

Access to the data portal

Info on APIs

Info for potential stakeholders

THE PAN-EUROPEAN VIRTUAL HUB DATA PORTAL
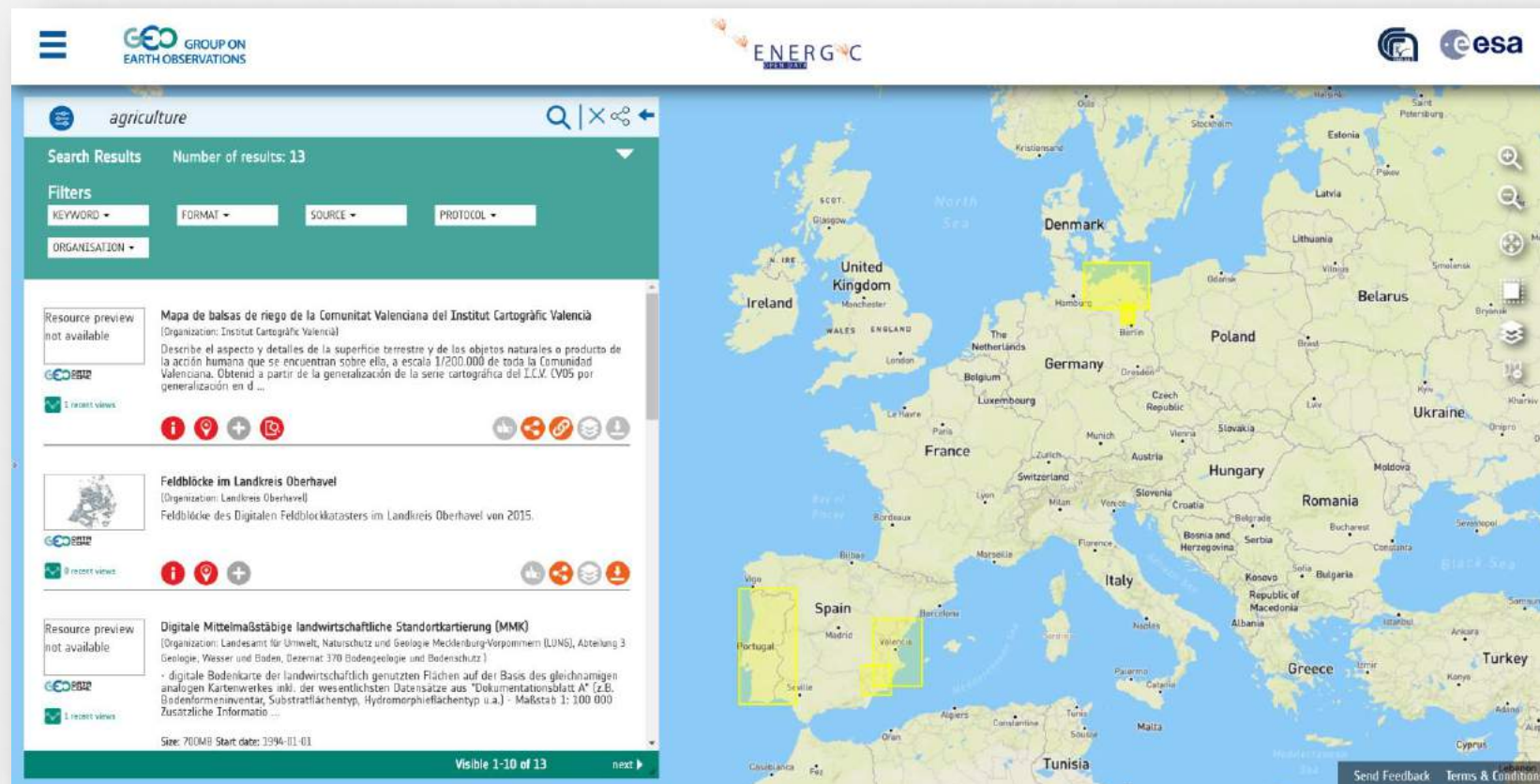
CONCLUSION

CONCLUSIONS 1/2

Virtual Hub

- An innovative solution for geospatial open data sharing has been proposed

- Addresses OD interoperability and governance issues implementing a **brokered System of Systems** architecture

- Addresses usability issues implementing **mediation,** dataset transformation for **harmonization,** and simple **APIs**

- **Lowers barriers** to participation in distributed systems for both users and resource providers
  - minimal cost impact on existing systems;

CONCLUSIONS 2/2

- **Accelerates interconnection** of disparate systems;

- **Facilitates sustainability**, reusability, extensibility, and flexibility of the infrastructure

- **Enhances multi-disciplinary interoperability** via introduction of new capabilities across multiple domains;

- **Removes need to impose common** (e.g. federal, "top-down") **specifications** and software components enabling a more adaptive "bottom-up" evolution of the infrastructure

Thank you for your attention!

LINKS

- Test portal for «Environmental analysis and Landscape Mapping» course

    http://vh-test.energic-od.eu/gi-cat/search

- Pan-European Virtual Hub

    http://www.geoportal.org/community/energic-od

- Noise capture APP

    http://noise-planet.org/