Copernicus Climate Data Store
Workshop ECMWF Reading, 3-6
March 2015

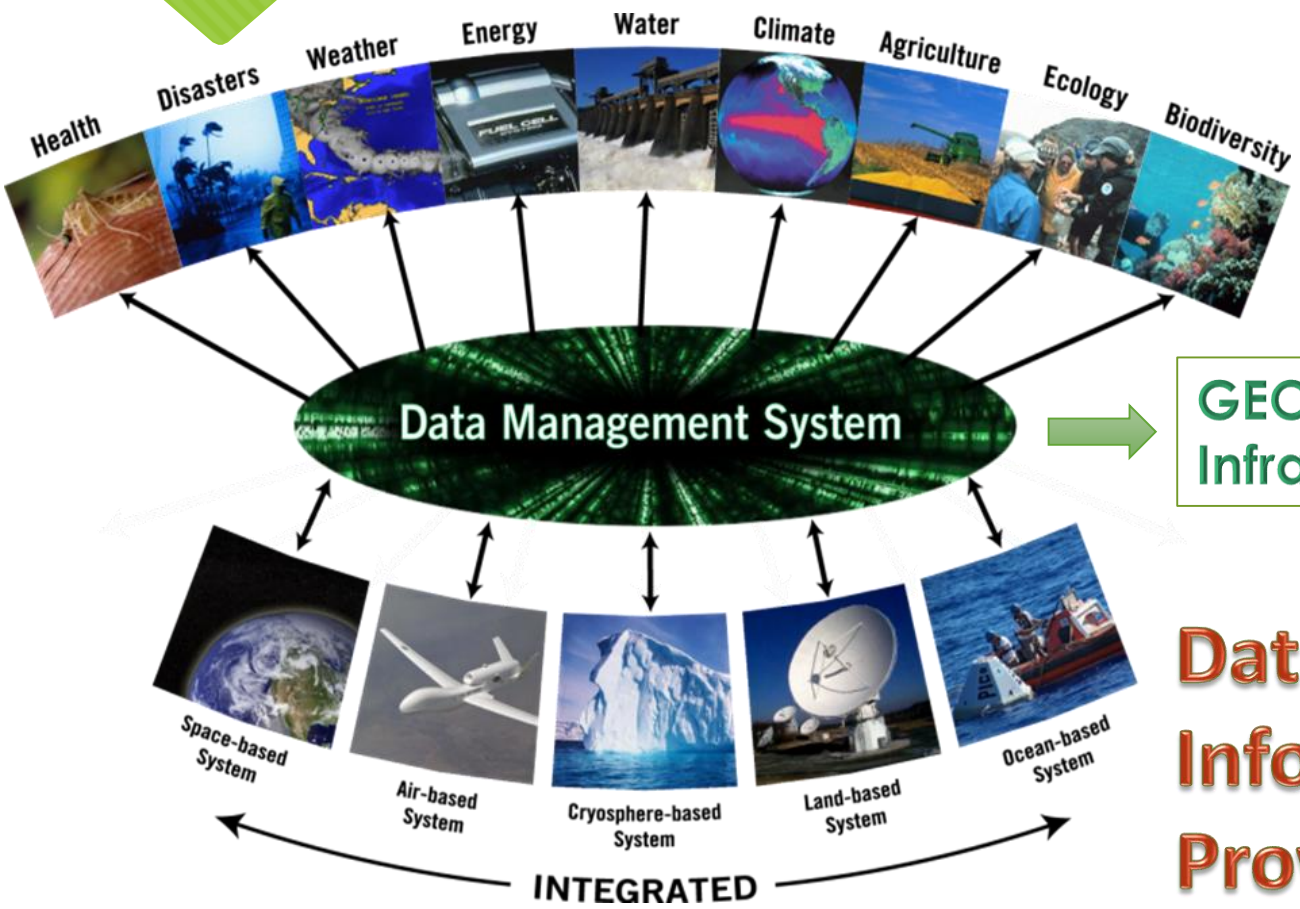
GEOSS Common Infrastructure (GCI) *and the* GEO Discovery and Access Broker (DAB)

GROUP ON EARTH OBSERVATIONS

Stefano Nativi

National Research Council of Italy (CNR-IIA)

Global Earth Observation Systems of Systems (GEOSS)



Science & Society

GEOSS Common Infrastructure (GCI)

Data & Information Providers

stefano.nativi@cnr.it

The GCI revolutions

►2005 ►2011 ►2013 ►2016

1st Revolution

2nd Revolution

3rd Revolution

4th Revolution

[Service driven]

[Brokering pattern]

[Big Data]

[… .]

… .

Virtual-Collaborative Management approach [Governance]

Collaborative-Acknowledged Management approach [Governance]

Users

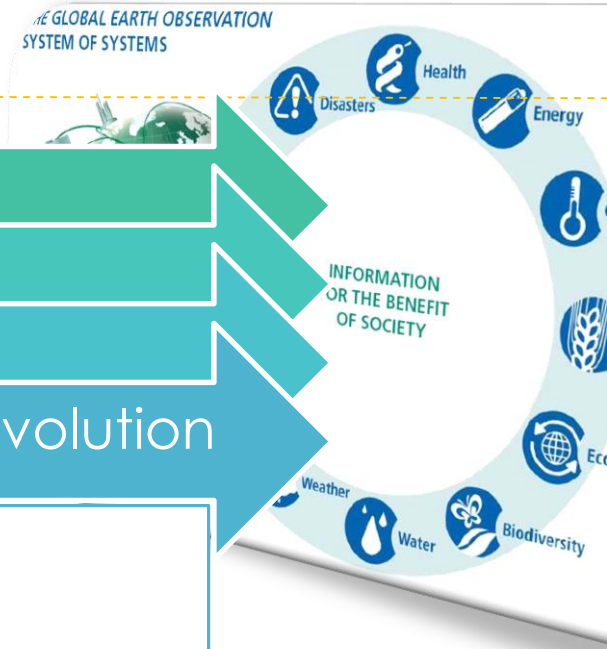
Providers

Social Ecosystem and build a Knowledge Base

Discovery [functionality]

Discovery and Access [functionality]

(re-)Use [functionality]



Service-driven collaborative Approach

stefano.nativi@cnr.it

Generally Recognized Barriers (especially for GEOSS)

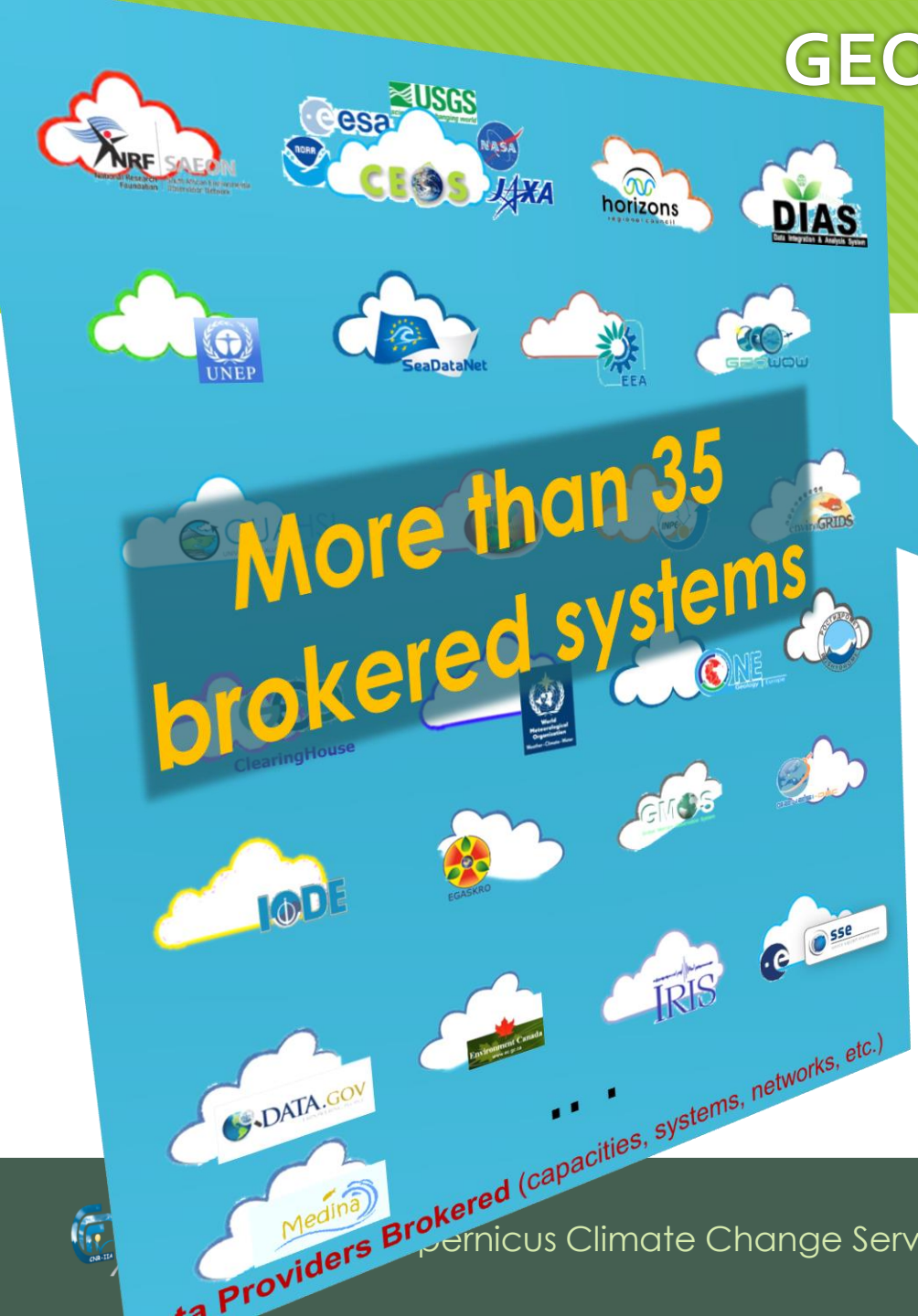
- **PROBLEM**: service **users need to know** the **nature** and **location** of **service providers**,
  - making it **difficult to bind** and **dynamically change** the **bindings** between users and providers
- **SOLUTION**: The **broker pattern** **separates users of services (clients) from providers of services (servers)** by inserting an intermediary, called a broker.
  - When a client needs a service, it queries a broker via a service interface. The broker then forwards the client's service request to a server, which processes the request

stefano.nativi@cnr.it

Collaborative-Acknowledged Management approach

stefano.nativi@cnr.it

GEOSS Information System



GEO Home Page

GEOSS Portal

Systems Registry

Semantic engines

stefano.nativi@cnr.it

GEOSS Assets (Nov 2014)

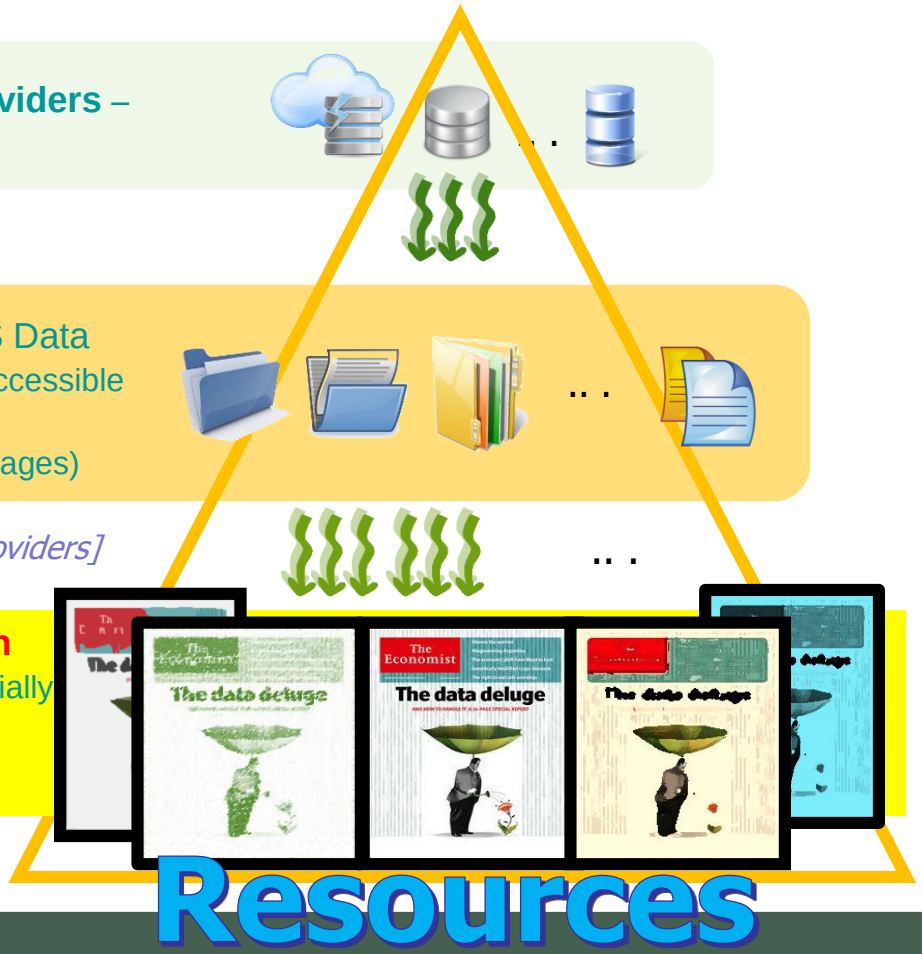
More than **35 brokered Data Providers** – capacities, systems, Communities

Publish

About 14 Million (2 Million GEOSS Data Core) Discoverable and potentially Accessible **first level resources** (mix of data collections, datasets and individual images)

Contain [source: data providers]

More than **82 Million (more than 50 Million GEOSS Data Core)** Discoverable and potentially Accessible **individual resources** (e.g. satellite scenes, rain gauge records)

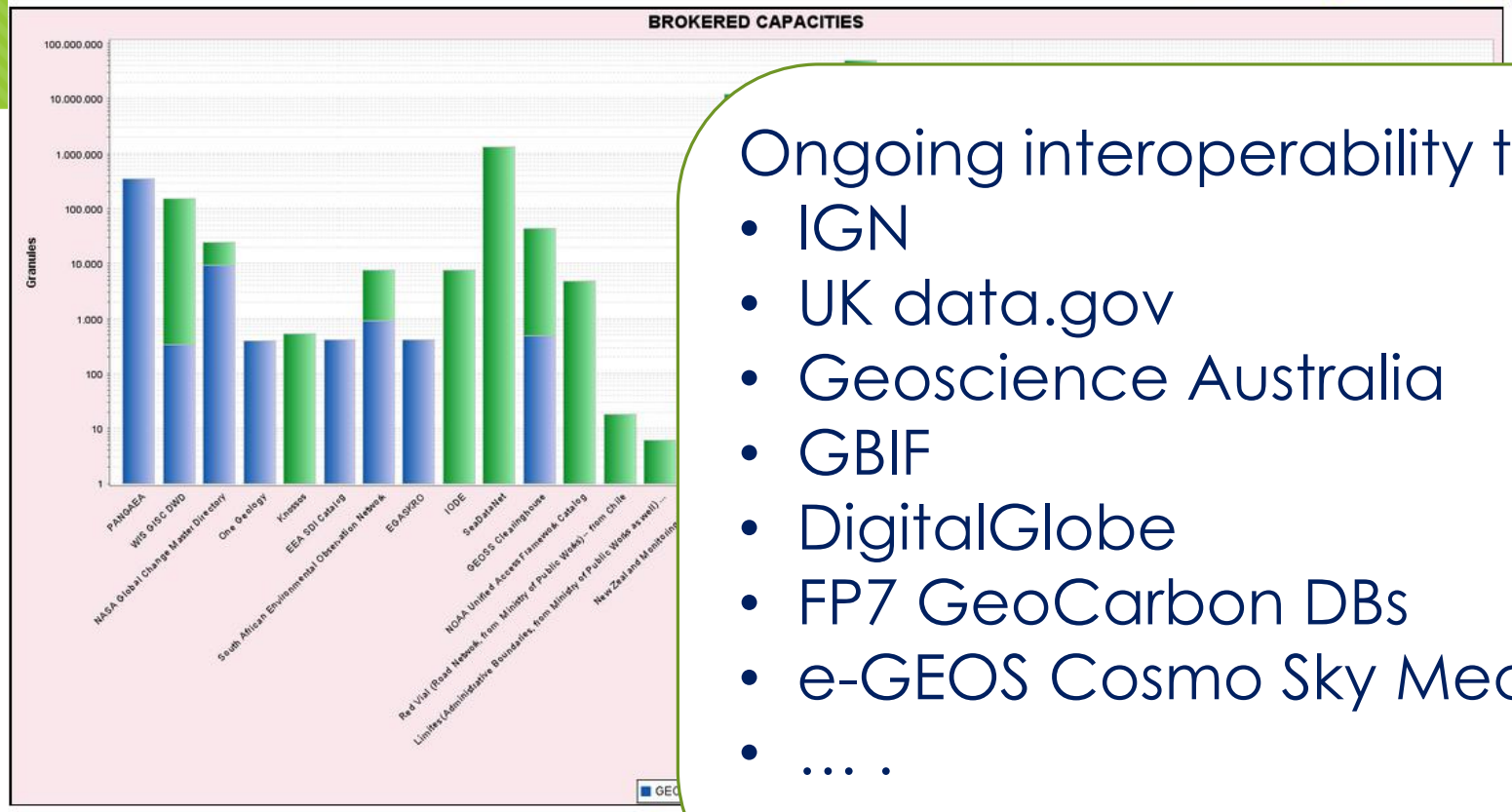


stefano.nativi@cnr.it

GROUP ON EARTH OBSERVATIONS

GEODAB
The GEOSS Discovery and Access Broker

Date 2014-07-14



Ongoing interoperability tests:

- IGN
- UK data.gov
- Geoscience Australia
- GBIF
- DigitalGlobe
- FP7 GeoCarbon DBs
- e-GEOS Cosmo Sky Med
- … .

	Total records	GEOSS Data Core records	Granules	GEOSS Core gra	
PANGAEA	349.718	349.718	349.718	349.718	
WIS GISC DWD	153.604	331	153.604	331	
NASA Global Change Master Directory	24.424	9.484	24.424	9.484	
One Geology	389	389	389	389	Number of records and granules harvested by GEODAB
GMOS Database	745	0	745	0	Number of records and granules harvested by GEODAB
ArcGIS Online ESRI	185.000	0	185.000	0	Declared by Provider
US Data Gov	75.968	0	75.968	0	Number of records and granules harvested by GEODAB
Total resources	13.537.677	1.199.572	71.690.212	51.197.805	

Published by

efano.nativi@cnr.it

Flexible, adaptable, and extensible interoperability Environment

Semantic

Access and Transformation

40 ma

56

ol Types

ations

Collaborative-Acknowledged Management approach

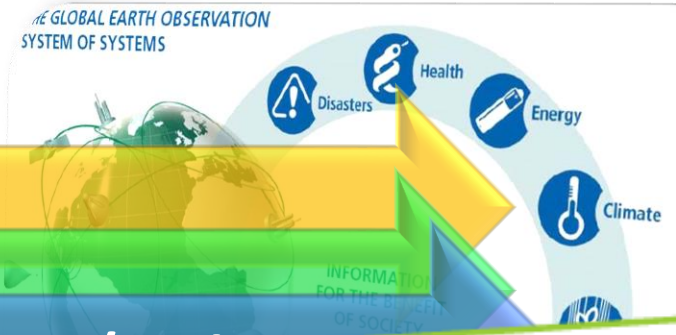
[Governance]

OGC WCS 1.0, 1.1, 1.1.2
OGC WMS 1.3.0, 1.1.1
OGC WFS 1.0.0
OGC WPS 1.0.0
OGC SOS 1.0.0
OGC CSW 2.0.2 Core, AP ISO 1.0, ebRIM/CIM, ebRIM/EO, CWIC
FLICKR
HDF
HMA CSW 2.0.2 ebRIM/CIM
GeoNetwork (versions 2.2.0 and 2.4.1) catalog service
Deegree (version 2.2) catalog service
ESRI ArcGIS Geoportal (version 10) catalog service
WAF Web Accessible Folders 1.0
FTP - File Transfer Protocol services populated with supported metadata
SITAD (Sistema Informativo Territoriale Ambientale Diffuso) accessor
GeoRSS 2.0
CERIF
KISTERS Web - Envi
DCAT
CKAN
HYRAX THREDDS S

stefano.nativi@cnr.it

Interoperability Standards

Interdisciplinary and Global Interoperability



Standardization

Federation

Common **Technological** and

interoperability

Organization

standards **mediation**

Global infrastructures (e.g. SoS, NoN)

Domain/discipline

Cohese infrastructures

Global infrastructures (e.g. SoS, NoN)

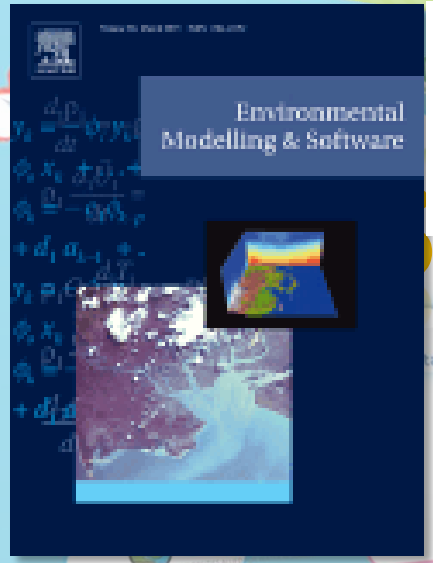
tef a i@cnr.it

Big Data challenges for GEOSS

- **Volume:** **Millions** of discoverable (small & medium) products; **Long** EO time/space series, … .
- **Variety:** Different **product types** (data, services, models, documents); **Data models**; **Protocols**; **Semantics**; **Granularity** levels; **Organizations**; **Maturity** level, … .
- **Veracity/Validity/Value:** **Evaluation** support, **Essential Variables**, Discovery **Ranking**, User **Feedbacks**, … .
- **Velocity:** serve Countries with **limited Internet access**, **Fast Discovery & Access**, Data Transformation and Analytics.
- **Visualization:** **Preview**, **Tiling** services, visual exploration, … .

Scalability and Reliability

stefano.nativi@cnr.it



Big Data Challenges in building the Global Earth Observation System of Systems

Stefano Nativi[1], Paolo Mazzetti[1], Mattia Santoro[1], Fabrizio Papeschi[1], Max Craglia[2], Osamu Ochiai[3]

n System

GEO
Home Page

S DATA, MORE INFO

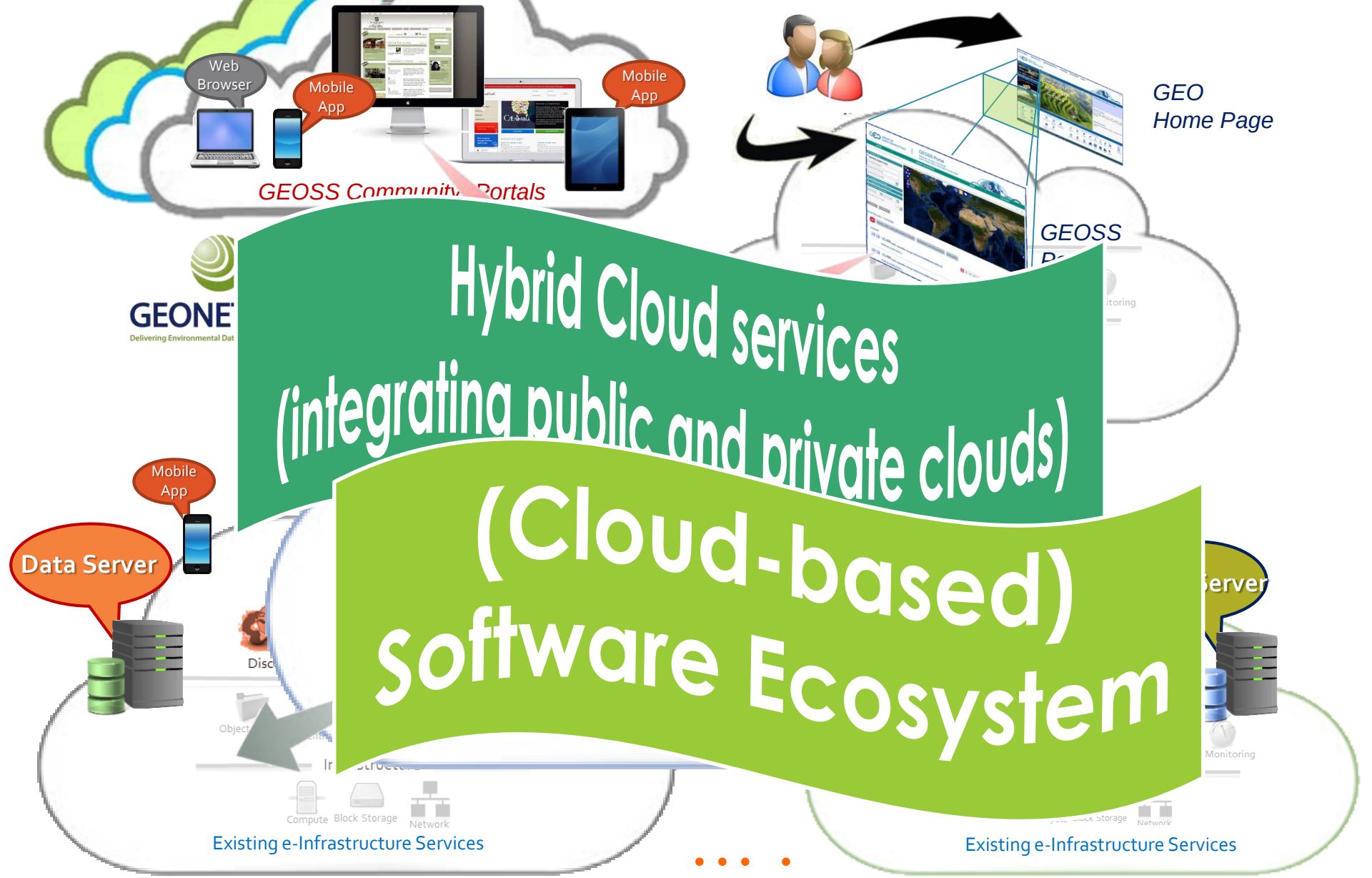
BLE CONFIGURATIONS

BROKERING APPROACH

CLOUD REVOLUTION

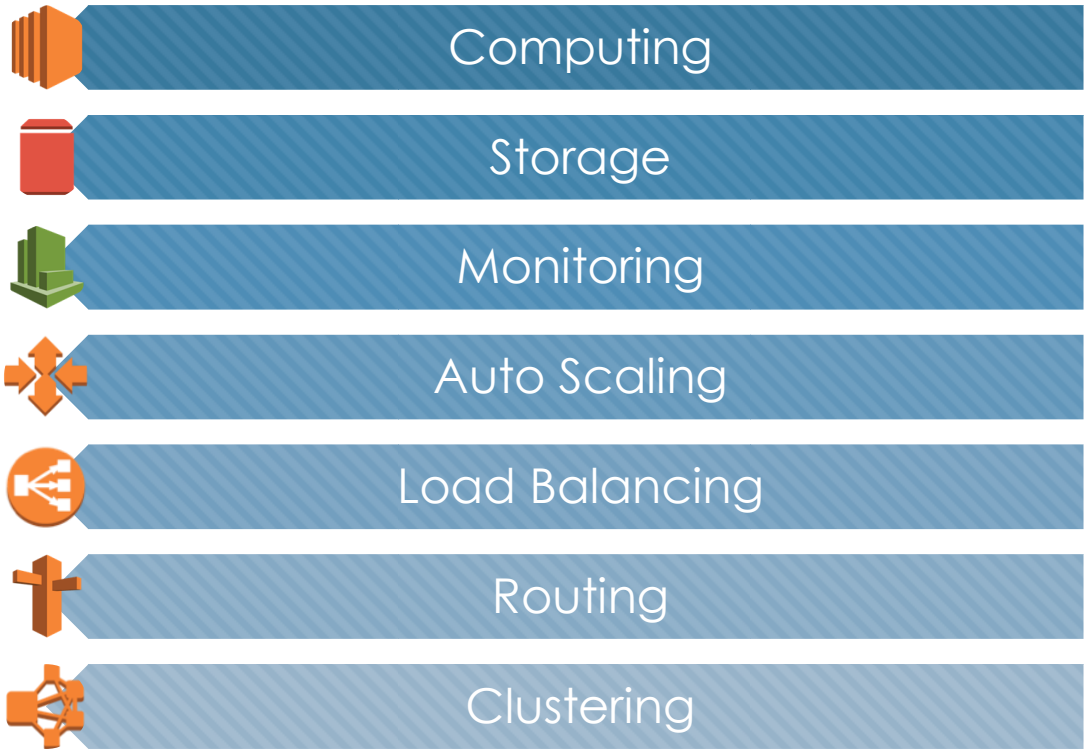
NOSQL DATABASES

stefano.nativi@cnr.it



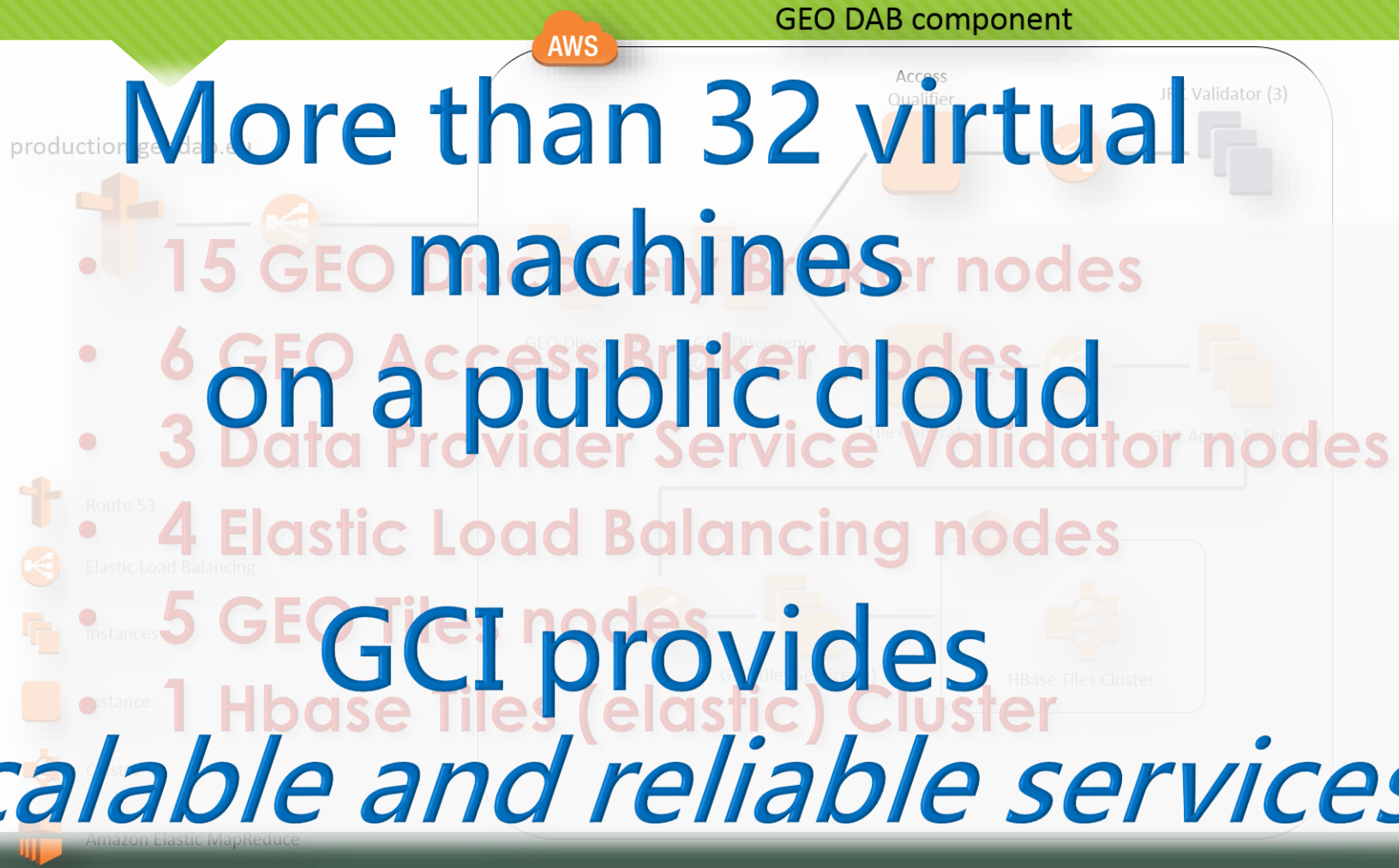
stefano.nativi@cnr.it

GCI (hybrid) IaaS and PaaS



stefano.nativi@cnr.it

GEO DAB in the Cloud



More than 32 virtual machines on a public cloud

GCI provides

Scalable and reliable services

stefano.nativi@cnr.it

Scalability benefits: Discovery Ranking Metrics

- A weight-based algorithm rewarding a set of criteria belonging to four high-level principles

Record score depends on **4 main Aspects**

Record **quality**

Data **accessibility**

Textual constraints Matching

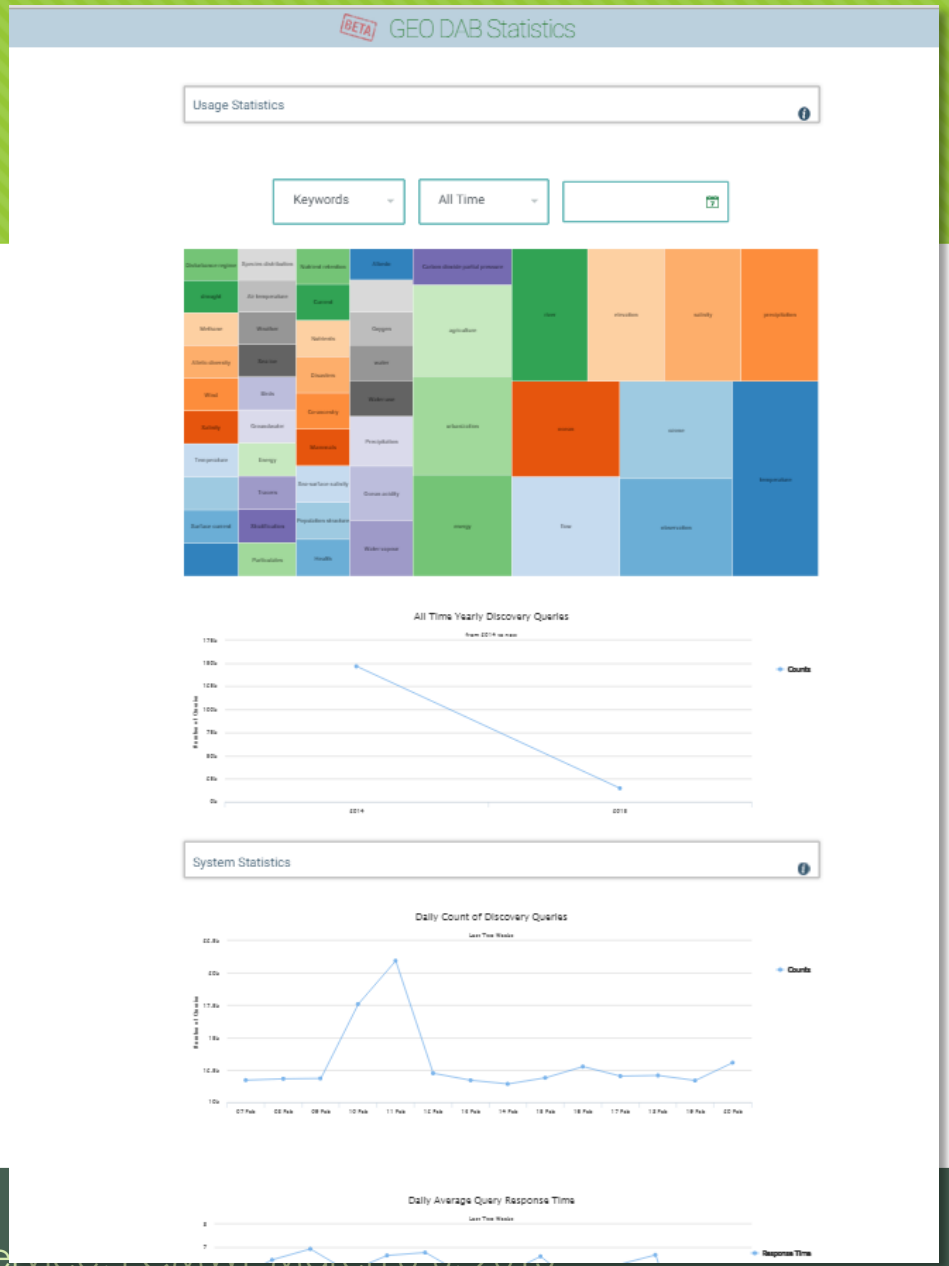
Domain constraints Matching

Essential Variables come first

Metrics is: Configurable and Flexible

stefano.nativi@cnr.it

DAB Monitoring and Statistics



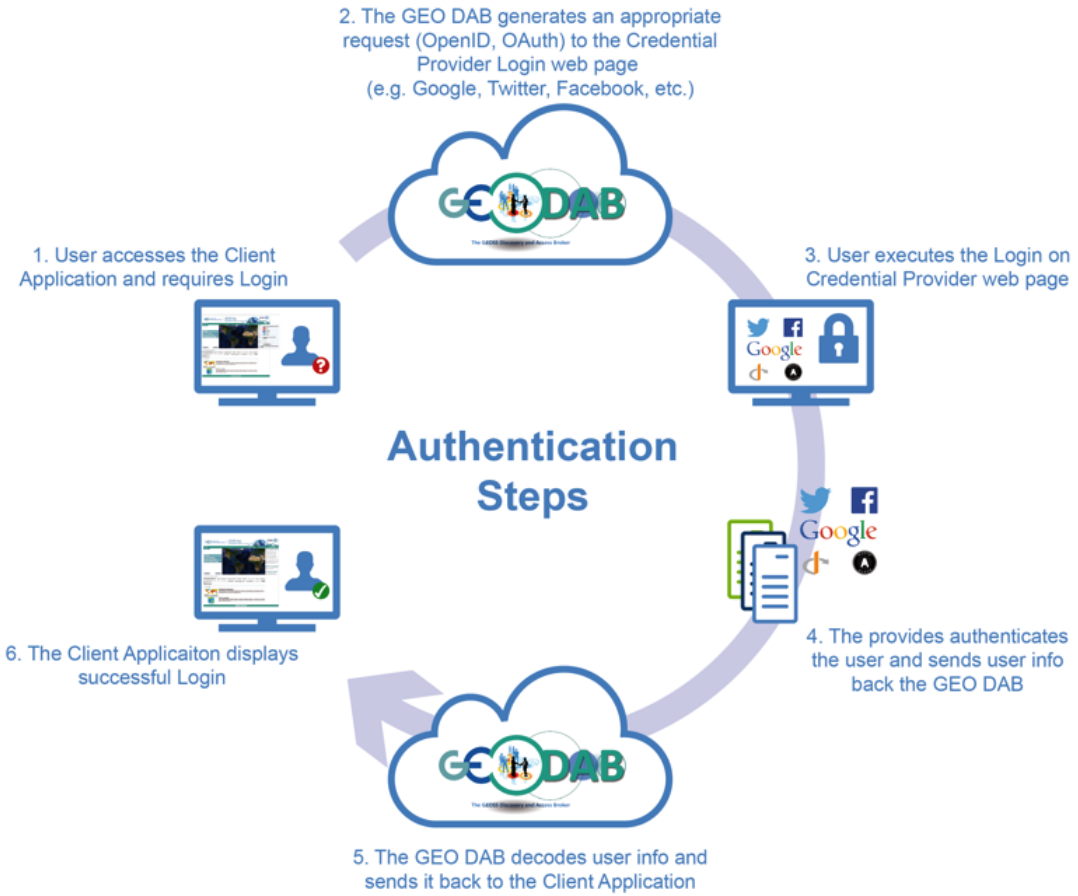
- Statistics
  - Year, month, week, day

- Most used **discovery constraints**
- Most used **remote data systems**
- Number of **discovery requests**
- Number of **access requests**

stefano.nativi@cnr.it

GEO DAB Authentication

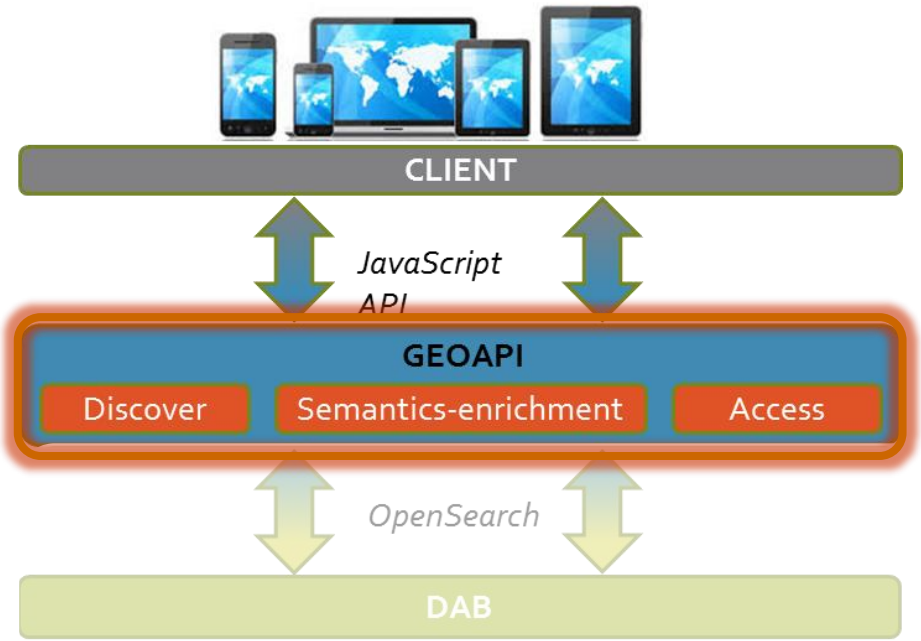
- Support existing and well-used users' credentials (e.g. **Facebook**, **Google**, **Twitter**, etc.)
- Support **OpenID** and **OAuth** protocols
- Support GEO DAB credentials



stefano.nativi@cnr.it

GEO Discovery and Access Broker APIs

- GEOAPI are **high level client-side Open APIs** (Application Program Interface)
- DAB users are typically **software agents**, such as **web-based or desktop client applications**



Designed and developed in **JavaScript** to simplify the development of applications and clients making use of the DAB

stefano.nativi@cnr.it

GCI and DAB Achievements

- (*From a "Catalog of Catalogs" to a*) **multi-disciplinary Brokering Platform**
- (*From discoverability to*) **accessibility and harmonization services**
- *(From an single infrastructure to a)* **cloud-based software ecosystem**
- *(From a virtual governance to a)* **collaborative-acknowledged governance**

stefano.nativi@cnr.it

Thank you !

Questions?

stefano.nativi@cnr.it