GROUP ON EARTH OBSERVATIONS

16th Workshop on meteorological operational systems
1-3 March, 2017

GEOSS Common Infrastructure and the Big Data challenges

*S. Nativi (1), J. Van Bemmelen (2), **M. Santoro** (1), G. Colangeli (2)*
O. Ochiai (3), P. De Salvo (3)

(1) Institute of Atmospheric Pollution Research,
National Research Council of Italy
(2) European Space Agency
(3) GEO Secretariat

GROUP ON EARTH OBSERVATIONS

Group on Earth Observation and Global Earth Observation system of systems

GEO AND GEOSS

The Group on Earth Observation (GEO)

GEO is a partnership of **more than 100 national governments** and in excess of 100 **Participating Organizations** that envisions a future where decisions and actions for the benefit of humankind are informed by coordinated, comprehensive and sustained Earth observations.

Ministers of the GEO member governments meet periodically to provide the political mandate and overall strategic direction for GEO. The Mexico City Ministerial Declaration from the GEO Ministerial Meeting in 2015 saw world leaders commit to support open Earth observation data for the next decade.

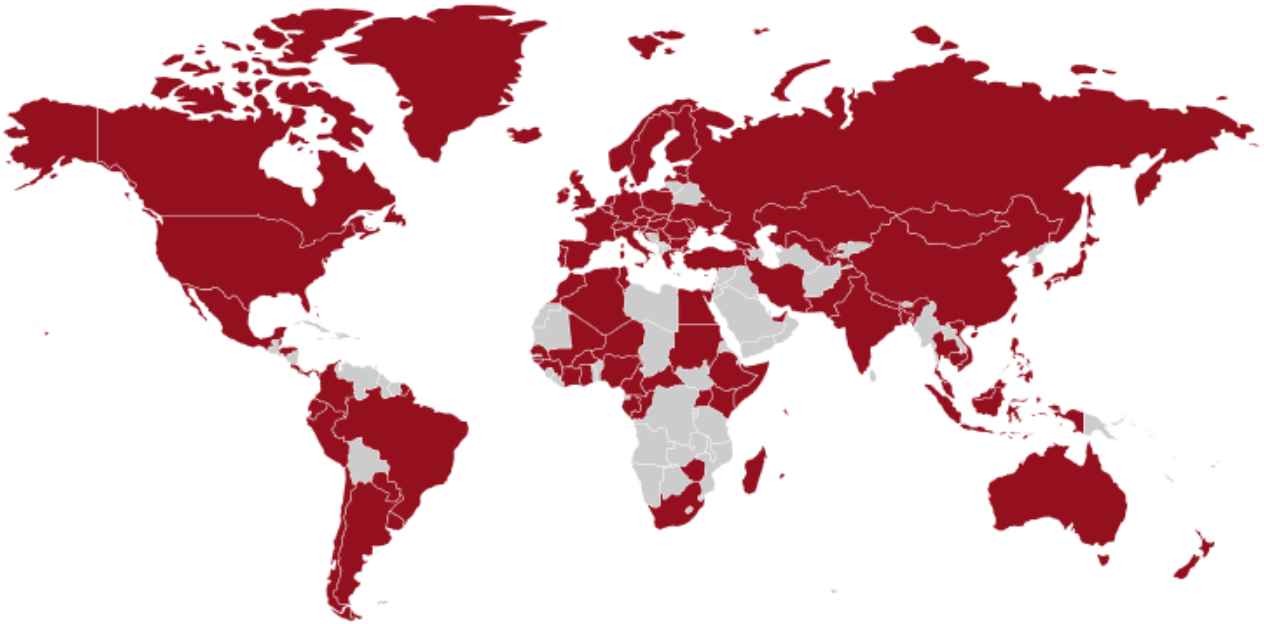
GEO is a unique global network connecting government institutions, academic and research institutions, data providers, businesses, engineers, scientists and experts to create innovative solutions to global challenges at a time of exponential data growth, human development and climate change that transcend national and disciplinary boundaries. The unprecedented global collaboration of experts helps identify gaps and reduce duplication in the areas of sustainable development and sound environmental management.

GROUP ON EARTH OBSERVATIONS

104 Member States

GEO Member Map for the year 2017

(Use slider under the map to change the year)

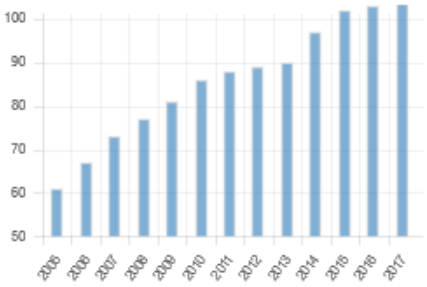


2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017

Number of Members (2017)

Africa:	27
Americas:	16
Asia/Oceania:	20
C.I.S.:	7
Europe:	34
Total:	104

Number of Members by year



GROUP ON EARTH OBSERVATIONS

106 Participating Organizations

Global Earth Observation System of Systems (GEOSS)

Together, the GEO community is creating a **Global Earth Observation System of Systems (GEOSS)**.

Earth observations from diverse sources, including satellite, airborne, in-situ platforms, and citizen observatories, when integrated together, provide powerful tools for understanding the past and present conditions of Earth systems, as well as the interplay between them.

GEOSS aims to better integrate observing systems and share data by connecting existing infrastructures.

There are more than 200 million open data resources in GEOSS from more than 150 national and regional providers such as NASA and ESA; international organizations such as WMO and the commercial sector such as Digital Globe.

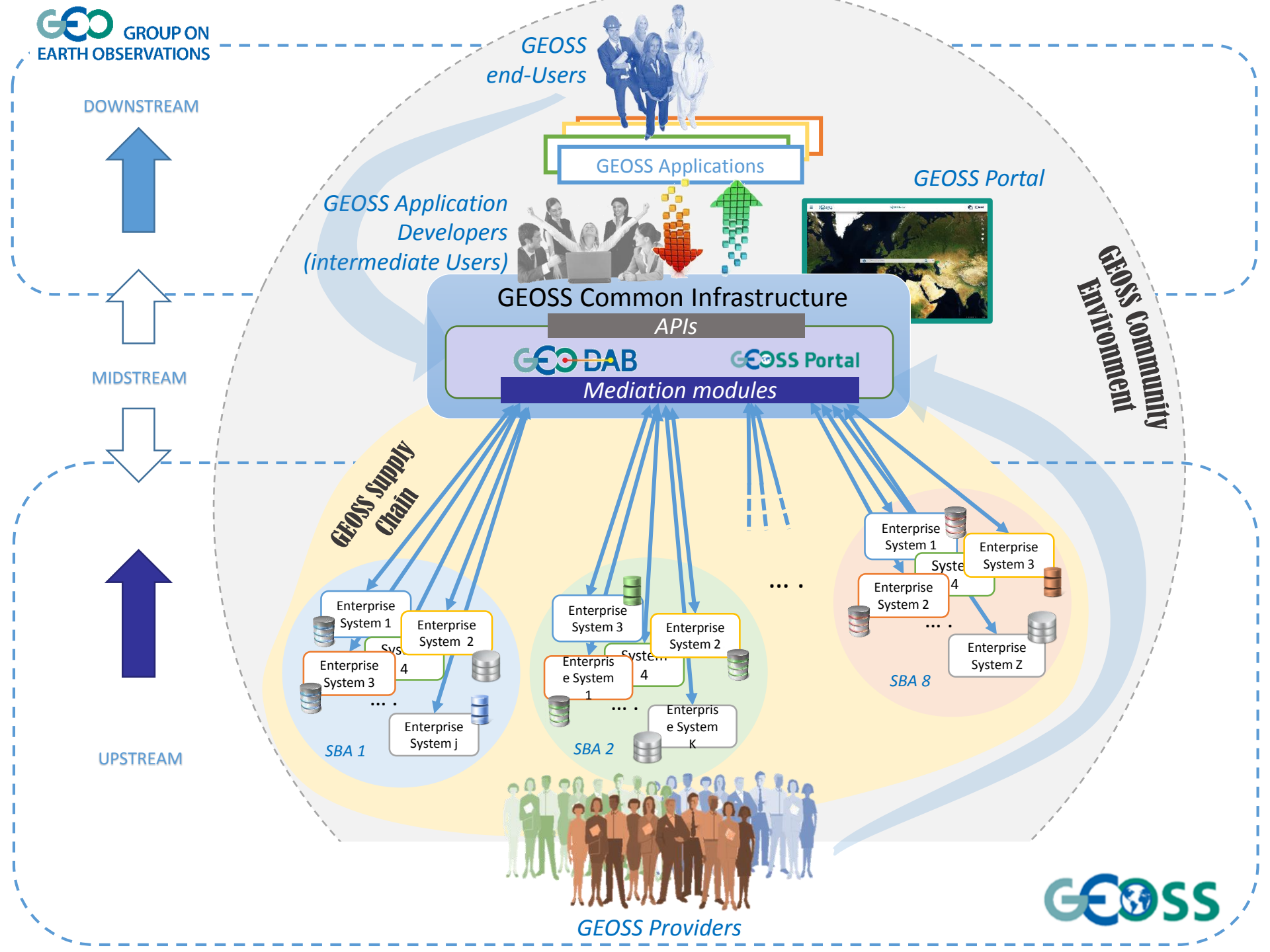
GROUP ON EARTH OBSERVATIONS

GEOSS Societal Benefit Areas

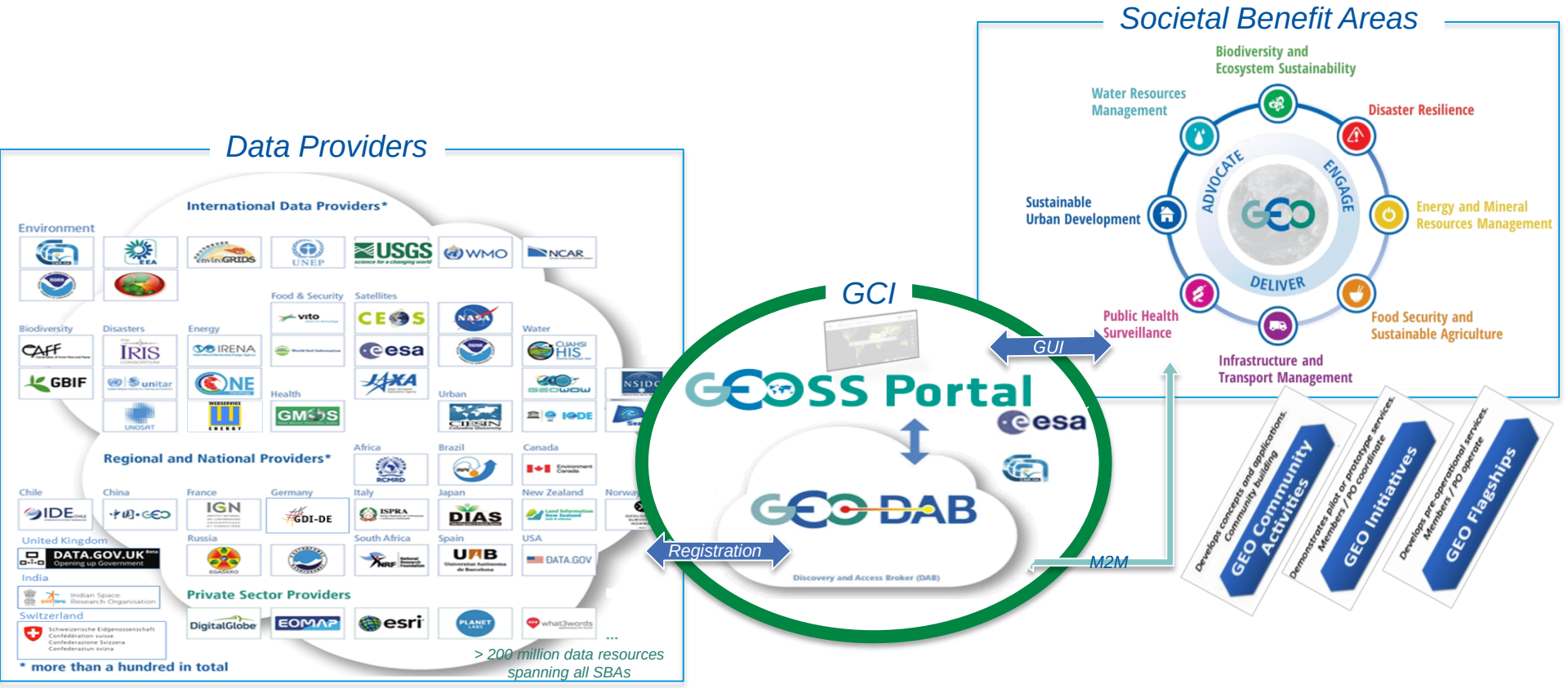
GROUP ON EARTH OBSERVATIONS

GEOSS Common Infrastructure (GCI)

IMPLEMENTING GEOSS


GROUP ON EARTH OBSERVATIONS
DOWNSTREAM
MIDSTREAM
UPSTREAM
GEOSS end-Users
GEOSS Applications
GEOSS Portal
GEOSS Application Developers (intermediate Users)
GEOSS Common Infrastructure
APIs
GEO DAB
GEOSS Portal
Mediation modules
GEOSS Community Environment
GEOSS Supply Chain
Enterprise System 1
Enterprise System 2
Enterprise System 3
4
Enterprise System j
SBA 1
Enterprise System 3
Enterprise System 2
Enterpris e System 1
4
Enterpris e System K
SBA 2
Enterprise System 1
Enterprise System 3
Enterprise System 2
Enterprise System Z
SBA 8
GEOSS Providers
GEOSS

GEOSS Common Infrastructure (GCI)



Enhanced GEOSS Portal - Overview

- Enhanced during 2016
- Accessible from *www.geoportal.org*
- Coordinated with *ESA*, *CNR-IIA*, *DG-RTD*, *DG-JRC* and *GeoSec*
- *Focus on **engagement, delivery and advocating***
- *Structured in **3 phases***
  - *1st phase – **2016**: interface restyling: completed*
  - *2nd phase – **2017/18**: deployment of major upgrades*
  - *3rd phase – **2019 onwards** – operations and evolutions*

GEO Discovery and Access Broker (DAB)

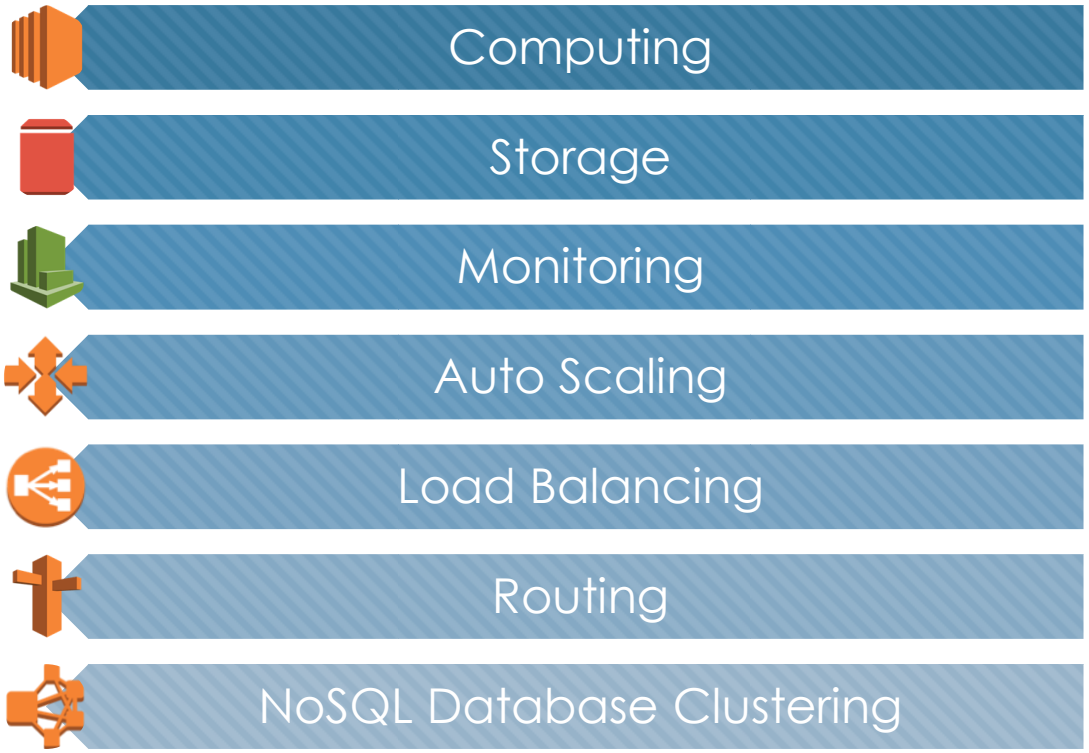
GEO DAB is a **brokering framework** that interconnects hundreds of heterogeneous and autonomous supply systems (the enterprise systems constituting the GEO metasystem) by providing mediation, harmonization and transformation capabilities.

GROUP ON
EARTH OBSERVATIONS

BIG DATA IN GEOSS

GROUP ON EARTH OBSERVATIONS

Big Data Enabling Technologies



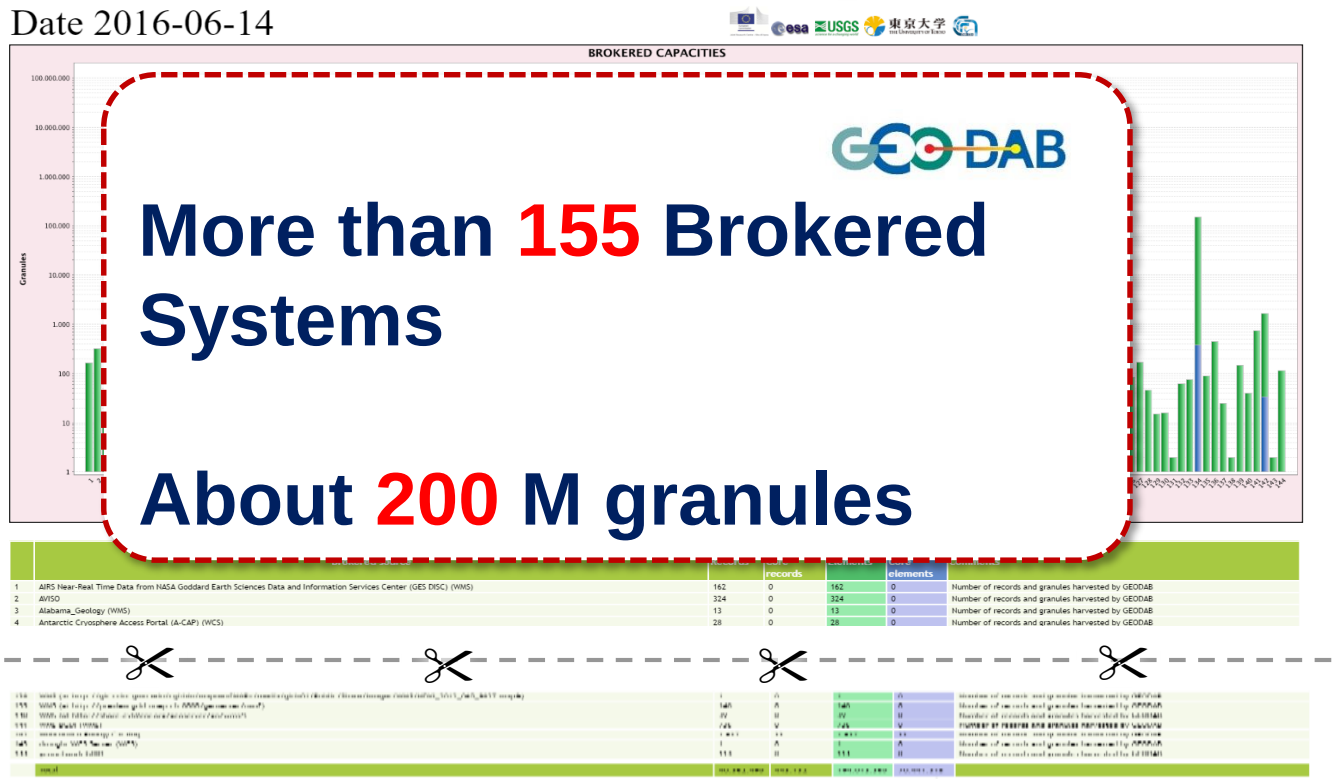
GROUP ON EARTH OBSERVATIONS

Big Data challenges for the GCI

VARIETY

Variety in GEOSS

- *Variety* is the most important *V* for GEOSS.



GROUP ON EARTH OBSERVATIONS

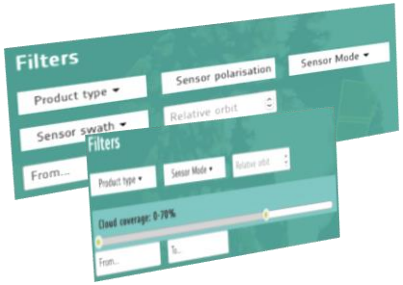
GEO DAB Supported Services

OGC CSW 2.0.2 AP ISO 1.0	INPE
OGC CSW 2.0.2 ebRIM EO	CKAN
OGC CSW 2.0.2 ebRIM CIM	DCAT
ESRI GEOPORTAL 10	GI-cat
OAI-PMH 2.0	ESRI GEOPORTAL 10
OpenSearch 1.1	NCML-OD
OpenSearch 1.1 ESIP	BCODMO
OpenSearch GENESI DR	NCML-CF
CKAN	NetCDF-CF 1.4
CUAHSI HIS-Central	FTP populated with supported metadata types
ESRI REST API 10.3	WAF Web Accessible Folders
OGC WCS	GeoNetwork (2.2.0 or greater)
OGC WMS	Ecological Markup Language 2.1.1
OGC WFS 1.0.0, 1.1.0, 2.0.0	NERRS (National Estuarine Research Reserve System)
OGC WMTS	HMA CSW 2.0.2 ebRIM/CIM
OGC SOS 1.0.0, 2.0.0, 2.0.0 Hydro Profile	HDF
OGC WPS 1.0.0	IADC DB (MySQL)
OGC CSW 2.0.0 Core	GrADS-DS
OGC CSW 2.0.2 AP ISO 1.0	FedEO
OGC CSW 2.0.2 ebRIM/EO AP	ARPA DB (based on Microsoft SQL)
OGC CSW 2.0.2 ebRIM/CIM AP	ESRI Map Server
IRIS Station	SHAPE files (FTP)
IRIS Event	KISTERS Web Environment of Canada
HYRAX THREDDS SERVER 1.9	Environment Canada Hydrometric data (FTP)
OAI-PMH 2.0 Harvesting	OpenSearch 1.1
GBIF	Earth Engine
DIF	RASAQM
HYDRO	EGASKRO
UNAVCO	SITADE (Sistema Informativo Territoriale Ambientale Diffuso)
CDI 1.04, 1.3, 1.4	File System
ISO19115-2	GDACS
THREDDS 1.0.1, 1.0.2	GeoRSS 2.0
THREDDS-NCISO 1.0.1, 1.0.2	Degree catalog service 2.2
THREDDS-NCISO-PLUS 1.0.1, 1.0.2	OpenSearch GENESI DR

Adopted Solutions – GEOSS Portal

- *User-centric, considering various user communities*:
  - GEO Flagships and Global initiatives
  - ESA Thematic Exploitation Platforms

- *SBA/Thematic Customization:*
  - ***Satellite****: includes smart filters for imagery (Landsat, Sentinel 2) and SAR-type (Sentinel 1) satellite data;*



  - ***Disater Resilience SBA****:* Earthquake events filters

GROUP ON EARTH OBSERVATIONS

Big Data challenges for the GCI

VOLUME

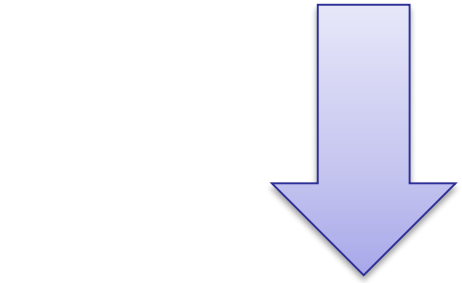
Volume in GEOSS

- GEOSS has to deal with the large amount of datasets provided by the end systems, e.g. millions of discoverable (small to medium size) products, and long EO time/space series.
- While GEOSS does not store the datasets, it has to collect metadata (at least for harvested catalogs) and provide effective discoverability.

Adopted Solutions

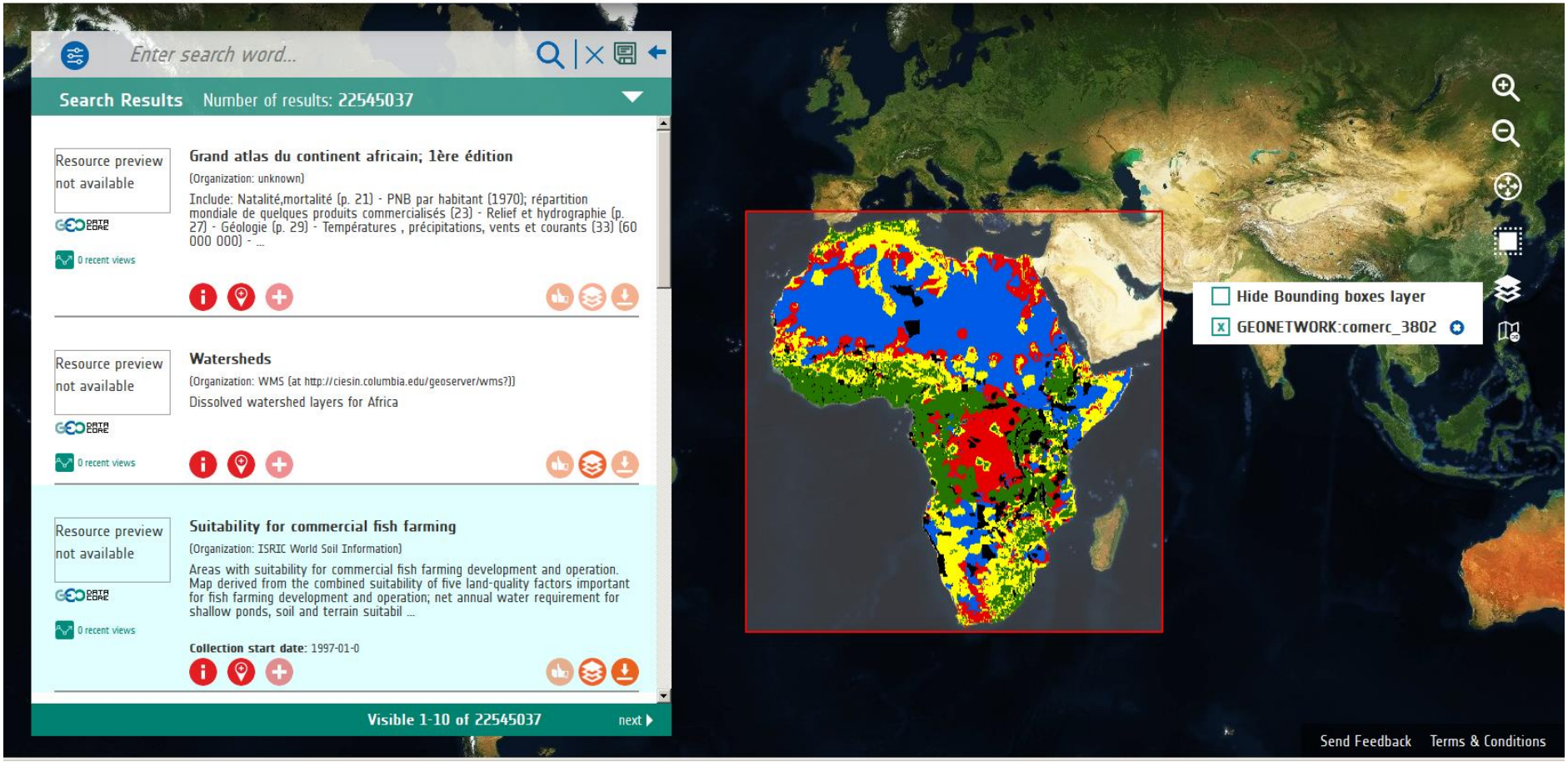
- Dealing with such numbers, normally constrained queries commonly match a large number of datasets.
- GCI addresses this challenge by returning a smaller and/or an ordered result sets.

Ranking and Paging



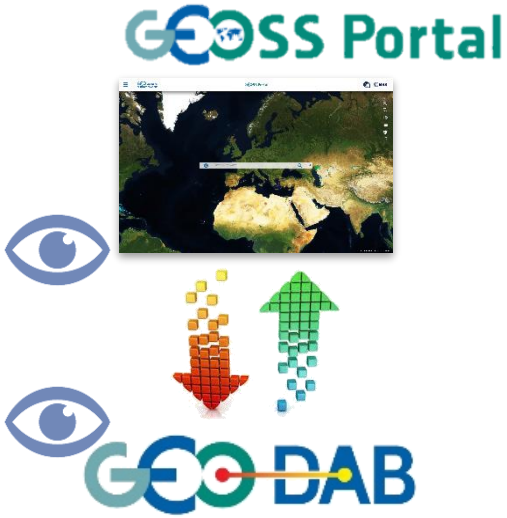
Views

Ranking and Paging



GEOSS View

- Definition:
  - Subset of the whole GEOSS resources defined by applying, via the DAB, a set of clauses
    - *Discovery clauses* (e.g. spatial envelope, keywords, sources, etc.)
    - *Access clauses* (e.g. data format, access protocol, CRS, etc.)
  - Defined “View” exposed on the GEOSS Portal



Consumer-defined View – i.e. Client-side These views are available only for the client application which defined the view.

Provider-defined View –i.e. Server-side These views are available for all client applications.

GROUP ON EARTH OBSERVATIONS

Big Data challenges for the GCI

VELOCITY

GROUP ON EARTH OBSERVATIONS

Velocity in GEOSS

- In GEOSS, *Velocity* related challenges include:
  - Processing rate to transform and preview data
  - Asynchronous approach for data access
  - Real-time (or near real-time) data access

Adopted Solutions – Fast Preview

- GEO DAB provides a fast preview service allowing to get data preview:
  - Metadata record is augmented by adding a reference to data preview; preview tiles at different zoom levels are generated in a batch mode.
  - To store and retrieve single tiles in an efficient way, GEO DAB utilizes a NoSQL key-value DB.
  - When available, GEO DAB utilizes data provider fast prview services by implementing the required mediation.
- GEOSS Portal uses allows Users to quickly evaluate discovered data before deciding the download.

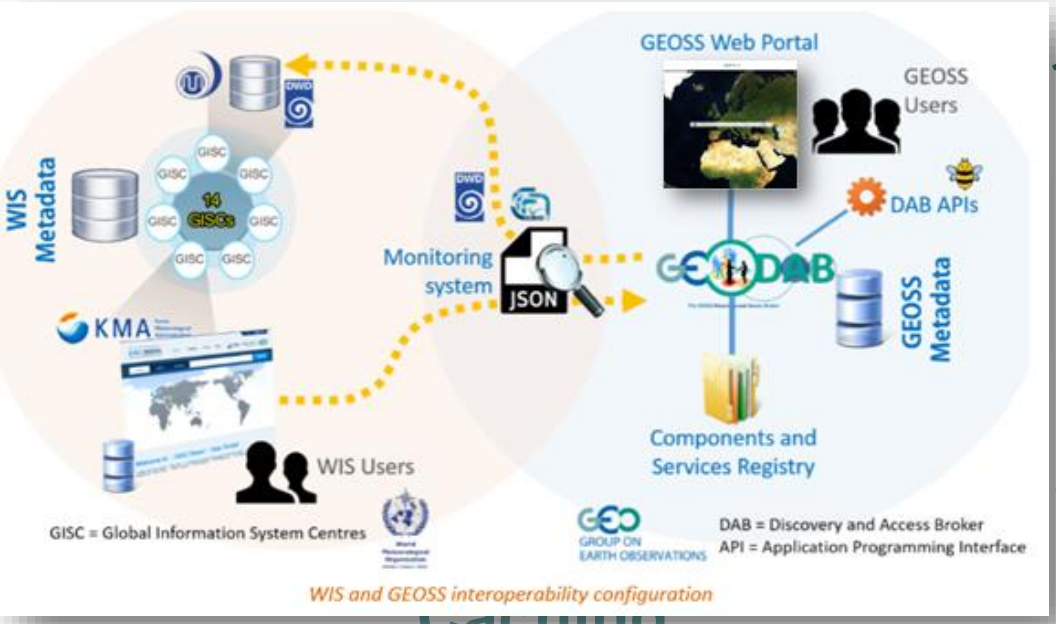
Adopted Solutions – Asyncronous Approach

- In an environment such as GEOSS … er which technique is implemented the … cases in which the required pr… …oo much time for a *cli…*
- The DAB + C… …ormation allows t… …ets according to a *co*… …nate Reference System, sp… …ent and resolution.
- Whe… …rmation workflow requires a long proces… …me, Users are allowed to opt for an asynchronous version of the same services.

This is still under testing with the Enhanced GEOSS Portal

Adopted Solutions – Real-time (or near real-time)

- GEOSS must support near real-time data discovery and access (i.e. GEOSS must be able to broker near real-time systems)
- Two strategies have been pursued to broker these systems:



Harvest information of near

Global Biodiversity Facility (GBIF)
INPE Steallite Imagery
ESRI ArcGIS Online
...

Caching

High-frequency Re-harvesting

GROUP ON EARTH OBSERVATIONS

Big Data challenges for the GCI

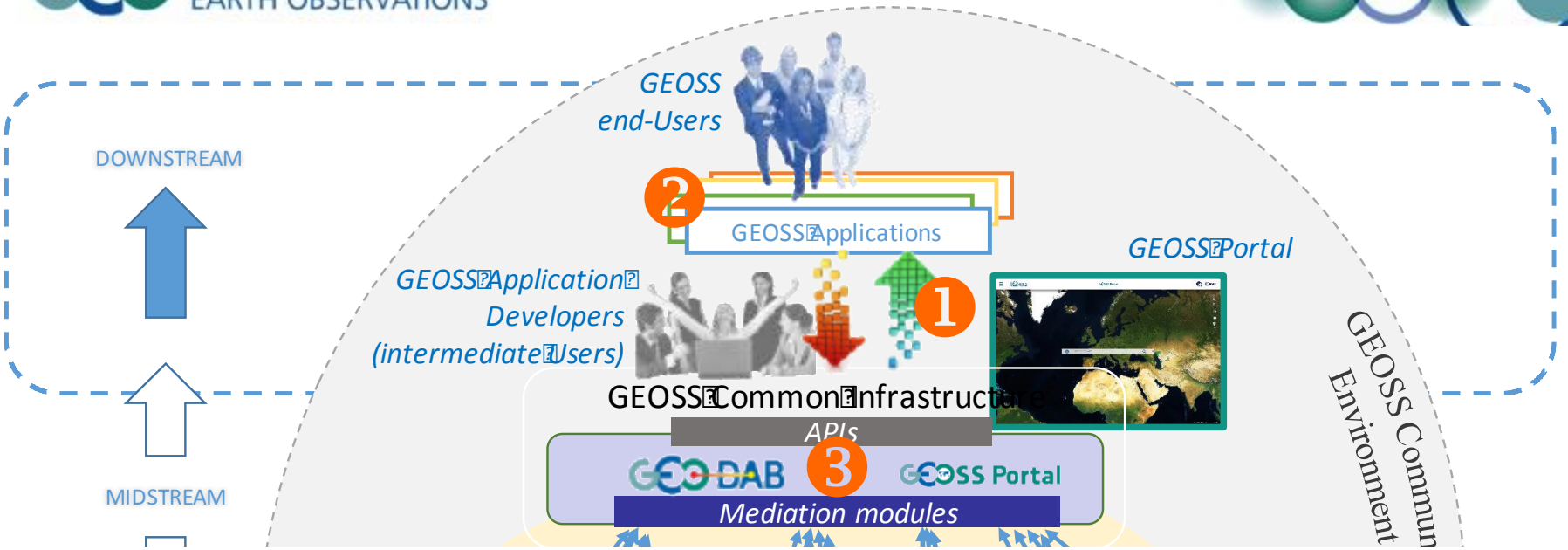
VISUALIZATION

Visualization in GEOSS

- In GEOSS, challenges related to *Visualization* stem from datasets heterogeneity and volume.
- In addition, GEOSS needs to address the requirement to support diverse (cross-)disciplinary applications targeting different Communities and User categories which have different needs, as for data visualization in an informative and significant way.

Adopted solutions

- GEOSS Portal customization:
  - In addition to what was described in *Variety* challenge, GEOSS Portal is focusing on providing ***resuable Portlets*** (for integration in external Community Applications) and ***custom visualization of results*** (e.g. display seismic events according to magnitude)
- A set of high-level APIs (Application Program Interfaces) have been designed and developed along with documentation and usage examples (the GEO DAB APIs) to allow the development of ad-hoc applications exploiting GEOSS content.



Different APIs for serving diverse Application development use cases (environments)

1

A set of standard ***Web service interfaces***:

- e.g. **OGC service interfaces**, **CKAN**, **OAI-PMH**, **FTP**, etc.

2

A set of APIs for software developers:

- *Client side* APIs:
  - (high-level) **JavaScript library**
  - … . (Python)

3

- *Server side* APIs:
  - **REST/JSON APIs**
  - **OpenSearch APIs**
  - **… .**

The GEOSS Discovery and Access Broker APIs

GROUP ON EARTH OBSERVATIONS

Big Data challenges for the GCI

VERACITY AND VALUE

Veracity and Value in GEOSS

- Giving access to a huge amount of datasets coming from different systems with their own mandate and governance, GEOSS has to consider the veracity and value of the published information.
- Particularly true if considering that GEOSS targets not only research communities, but also decision and policy makers, and therefore the veracity and value of the pub- lished information may affect relevant decisions.

Adopted Solutions

- GEOSS Data Mangement Working Group provides a a set of **Data Management Principles**, including quality-related aspects;

- **Essential Variables:**
  - EVs can be defined as those parameters required for study, reporting, and management of problems in a specific scientific or societal domains.
  - This effort is particularly important for an infrastructure such as the GCI: the formalization and use of the EVs concept, and related instances, allows extracting the most valuable data matching User's request.

GROUP ON EARTH OBSERVATIONS

Conclusions

- In the past 10 years GEOSS has developed a truly Global and multidisciplinary System-of Systems
- A valuable framework to experiment and learn how to face Big Data challenges –in particular Variaty and Volume ones.
- The new GEOSS Portal + DAB platform signifcantly improved the discoverability and accessibility of sahred GEOSS resources, addressing more and more User requirements.

GROUP ON
EARTH OBSERVATIONS

Thank you

GROUP ON EARTH OBSERVATIONS

Backup