



# **ESA Earth Observation Big Data R&D Past, Present, & Future Activities**

**CEOS WGISS Meeting #37  
14-18 April, 2014  
Cocoa Beach, Florida, USA  
Hosted by NASA**

**M.Albani, S.Loekken, J.Farres (ESA)**

1. Opportunities and Challenges; High Level Objectives
2. Past and present activities
3. Lessons learnt
4. Future activities



Definition of BIG DATA includes:

1. The **size** (volume) of the data sets
2. The degree of **diversity** and **complexity** within the data sets
3. The *amount of value* that can be derived from innovative analysis techniques applied to diverse and complex data sets

Aspects relevant for Earth Observation:

- A. Increasing **amount** of **EO space data** (past, current and new satellite data)
- B. Increasing **diversity** of **EO space data** (in media, formats and processing)
- C. Increasing **amount and diversity** of **non-space data** used in conjunction with EO space data (e.g. ground GPS measurements, marine buoys measurements, atmospheric plane measurements, etc...)

→ Hence ESA Earth Observation Directorate *strong* interest in Big Data

# ESA EO Big Data R&D Opportunities and Challenges



R&D agenda driven by opportunities and challenges of Big EO Data

**Opportunities** - Sentinels, Earth Explorers, Copernicus Contributing Missions, TPMs, long-term EO archives (e.g. ERS, Envisat and others), in-situ networks & models provide:

- ✓ an unprecedented **precision and variety** of global data on the state of our planet
- ✓ an **unprecedented opportunity for insight** into how our oceans, atmosphere, land and ice operate and interact as part of an interconnected Earth System
- ☒ Opportunities for science, for commercial exploitation, applications, & ICT

**Challenges** – technical but not only

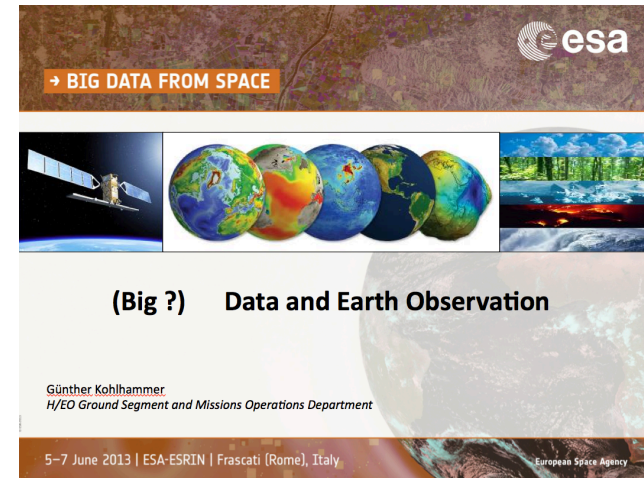
- ✓ Technical challenges related to increasing Volume, Variety and Velocity of data
- ✓ ICT provisioning – scalability, high volumes & dissemination peaks, processing bursting etc.
- ✓ Federation of European ICT and Ground Segment capabilities – leverage existing initiatives
- ✓ GS ops concept evolution – Collaborative platforms, resource pooling, processing to the data
- ✓ Sustainability – cost savings (also ICT), funding & payment models

# ESA EO Thematic Exploitation Platforms

## ESA EO Big Data Workshop Outcome



- From the conclusions of the **ESA Big Data Workshop 2013**:
- "The increasing demand for and usage of large amount of EO data in science and exploitation*
- opens big **opportunities** for research
  - brings necessarily **new mechanisms and schemes for cooperation**
  - requires **new approaches, services, and ops concepts** for the PDGS (e.g. exploitation platforms etc)
  - requires **good technical, programmatic and industrial coordination"**



### Corresponding **ESA EO objectives**:

- Further increase the amount of data being placed online for the public to use and develop online services allowing patrons to access, process & interact with space data in synergy with non-space data
- Continue **evolution of EO ground segment and mission operation strategy** in particular focusing on new approaches and technologies (Big Data) in collaboration with Member States and Industry
- **Facilitate effective and economically sustainable use of modern computing infrastructures and ground segment capabilities by European industry**

- ☒ Elements of Big Data handling and exploitation in EO domains at a maturity level where R&D and **coordinated action** is required. And not only for technology.

1. Opportunities and Challenges; High Level Objectives
2. Past and present activities
3. Lessons learnt
4. Future activities

# ESA EO Big Data R&D Past and Present Activities



ESA EO has pursued (amongst others) the following activities and initiatives:

1. **ESA Research and Service Support (RSS)**

- Support EO community in exploiting EO data and researchers and service providers in the development of applications for value added information

2. Activities addressing **ICT provisioning, Grid and Cloud computing, IaaS**

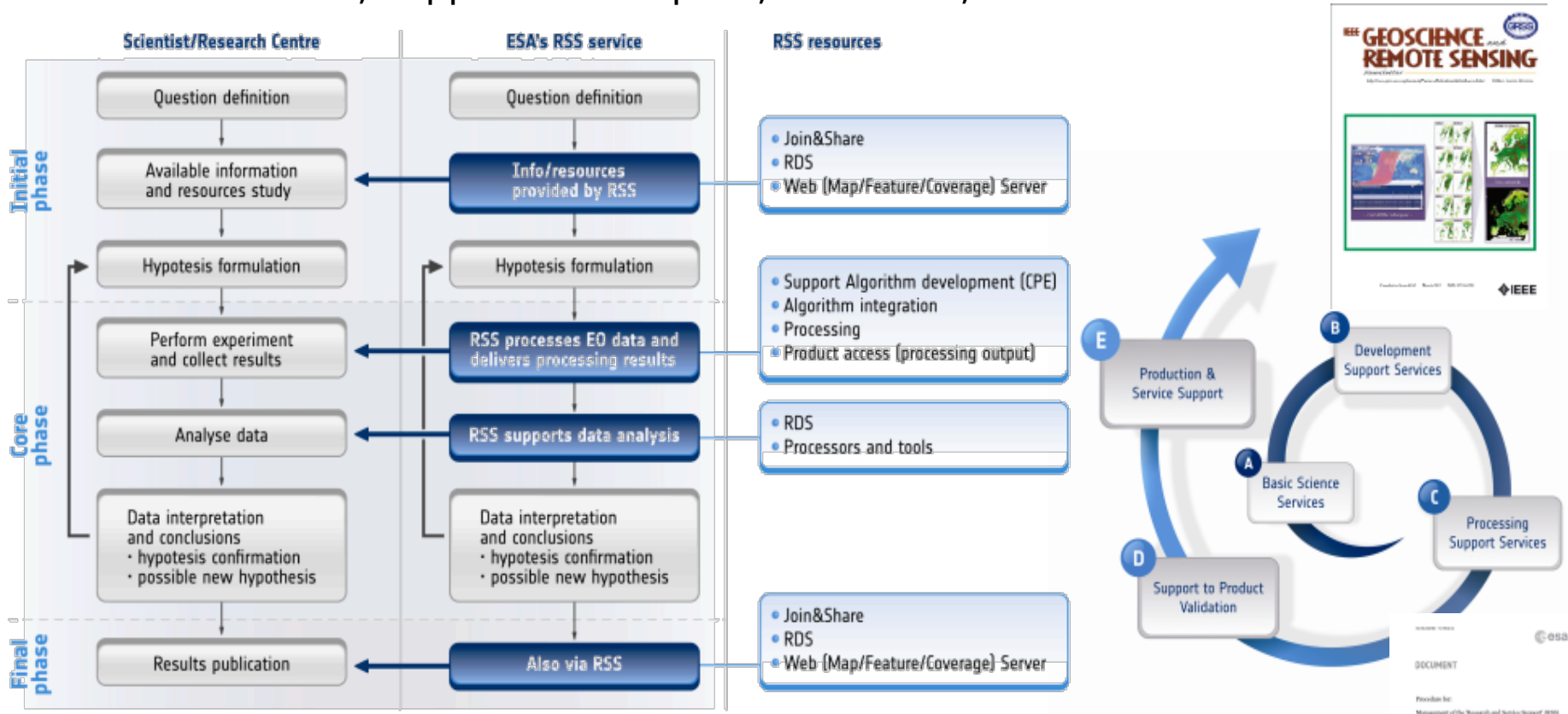
- Grid computing on demand – G-POD (since 2004)
- Outsourced services for data hosting (since 2010)
- EO data re-processing (on Amazon) (in 2011)
- Dissemination and processing (on Hetzner, in 2010)

3. **Complimentary GS operations concepts** – User Exploitation Platforms; intersecting Big Data exploitation, Ground segment capabilities, and ICT

- Geohazard SuperSites Exploitation platform (with Helix Nebula)
- Exploitation Platform 4 Soil Moisture (with CEMS)
- **Thematic Exploitation Platforms (TEPs)** initiative

# ESA EO Big Data R&D Past and Present Activities: RSS

Support to PIs, services and projects, technology pilots and testbeds, standardisation, support to eduspace, 3D shows, more...



<http://rssportal.esa.int/rss-portal.php>



### Exploitation Platforms: **Evolution of complimentary ground segment operations concepts**

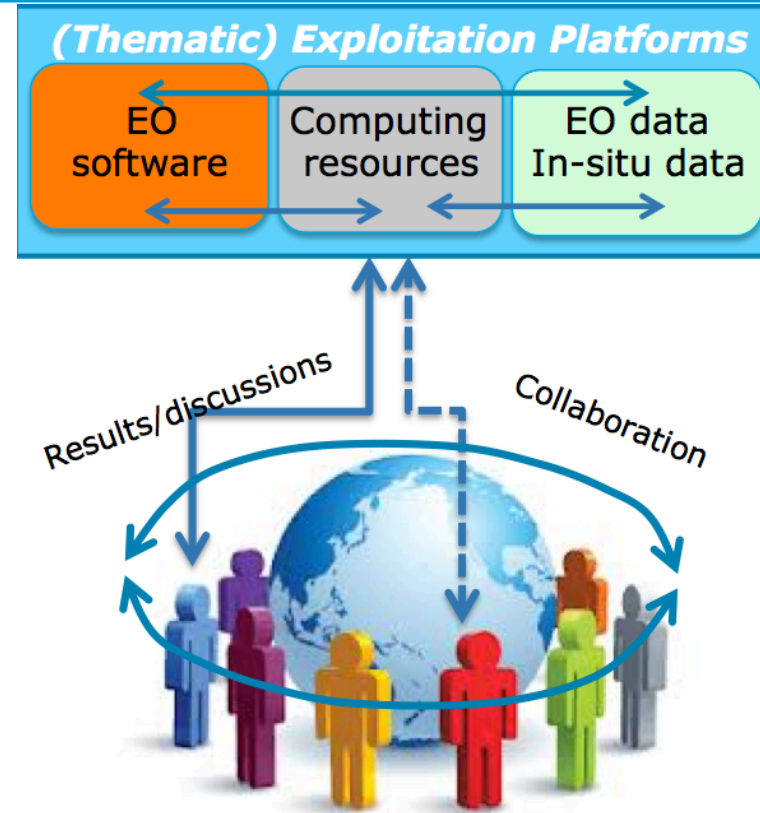
Implement an ecosystem of 5 Thematic Exploitation Platforms in the 2014-2016 time frame

**Request for Information (RFI)** issued September 2013:

- Gauge interest and gather ideas from member states, industry, academia, & user communities
- Gather information regarding potential themes and corresponding activities to be used as input to the preparation of a formal ESA Invitation To Tender.

**Response** far beyond expectations

- 48 proposals in a 1000-page corpus, from all ESA member states, covering 17 distinct themes
- Detailing possible Thematic Exploitation platforms
- Rich source of information also regarding initiatives intersecting Big Data exploitation, Ground Segment capabilities, and ICT in Europe



TEP = **Virtual Workplace** bringing together a **Community, relevant data, tools, and adequate computing resources on an open and collaborative framework around a thematic domain**

☒ Bring processing to the data

☒ The 'place' where data exploitation is done

1. Opportunities and Challenges; High Level Objectives
2. Past and present activities
3. Lessons learnt
4. Future activities

# ESA EO Big Data R&D

## Lessons Learnt: Cloud (IaaS) ICT



### Cloud ICT Provisioning

1. As soon as ICT needs can be predicted and planned, we found IaaS more expensive than rental or dedicated hosting solutions
2. Flexibility of Public IaaS less appealing when internal resources are pooled, virtualized and managed as an internal cloud.
3. But: IaaS services allow to **size down internal ICT resources** to the “fixed” need and ensure their maximum utilization; e.g. using external provisioning for the “variable” need

### ▸ Hybrid ICT provisioning

### IPR

1. Licensing schemes for virtualized/IaaS multi-user often not in place

### ☒ Evolve with data and IPR providers

### Cloud Service Levels

1. Terms/Conditions in Public Clouds expressed surprisingly low commitments
2. Cloud opportunities **may** become risks when applied to critical systems

### ▸ Develop multi-sourcing

### ▸ Plan contingency scenarios for services hosted in Public Clouds

### Application Areas

1. Dissemination and on-demand processing  $\leq$  because needs are variable and depending on user demand
2. Secondary archive and re-processing  $\leq$  because needs are limited in time
3. Temporary resources for integration, testing and demonstration  $\leq$  because needs are limited in time
4. System sizing  $\leq$  because needs are unknown

### Important areas where remote sensing services can gain from Cloud/IaaS Computing

### User Expectations

1. Open Data: All data discoverable, accessible online & free; arranged in long time series of coherent data from different providers
2. Open Computing: Users are able to perform processing directly on the cloud using virtual servers. Users can choose their preferred cloud provider
3. Open Source Software: All basic/platform software is open and freely available; Applications can be easily ported across clouds
4. Open Collaboration: Data and applications can be easily shared with other users
5. Managed services: Expert support to data exploitation through managed services

**=> Be up to users' expectations**

## From the TEPs RFI: Maturity of Big Data exploitation in Europe

- Several **initiatives underway** intersecting Big Data exploitation, GS capabilities, ICT
- Initiatives (and funding) predominantly national, but not coordinated – **federation** of existing capabilities emerges as most frequent request
- Exploitation initiatives target national ICT infrastructures and GS capabilities where these are available. **Explicit requests for help/direction** where they're not
- Operational providers of EO based information perceive as **most imminent problem** the implementation of infrastructure and data access for customized high volume service delivery
- Even mature areas are forced to spend significant effort on developing **heavy processing capabilities** (rather than on data analytics)
- Proponents look for synergies with other programme contexts (e.g. H2020)

- ☒ Maturity of Big **EO** Data exploitation initiatives in Europe still lagging – science, applications, value added, & ICT struggle with defining an approach
- ☒ Sustainability is problematic and often related to access to adequate ICT infrastructure and data at manageable cost – SMEs perceive a significant competitive disadvantage
- ☒ Federation and coordination (technical & programmatic) among existing and planned initiatives is a must

1. Opportunities and Challenges; High Level Objectives
2. Past and present activities
3. Lessons learnt
4. Future activities

Research and Service Support **managed service** crucial to help users in EO data exploitation

 Must evolve as the user needs evolve – also addressing Big Data

1. Increased scope of available data and services
  - For ERS, ENVISAT, Earth Explorers, Sentinels, TPMs, and some EUMETSAT missions
  - PIs support, extended time series, multi-sensor data fusion, cal/val groups, Instrument Processing Facility Test bed
2. Federated service and infrastructure
  - Virtual Data Farm with increased storage and processing capabilities
  - Additional (Federated) Resources, including Supersites (e.g. Geohazard); Academic Cloud; Collaborative GSs; Thematic Exploitation Platforms; and Commercial Clouds, ...

Following the TEPs RFI - in the planning horizon of 2014-2016, **implement an ecosystem of 5 Thematic Exploitation Platforms**

☒ **Coastal, Forestry, Hydrology, Polar, and Urban** TEPs up to end 2016

- ✓ **Thematic** – addressing thematic domains with many possible applications
  - ✓ **Science** – aimed primarily (but not only) at scientific exploitation of EO data
  - ✓ **Exploitation platform** – use cases for full collaborative work environment – Data exploitation, Product development, New service development etc.
  - ✓ **Harmonized** – based on *common components* and *standards* for generic platform capabilities (such as user management or data discovery), + domain-specific components (such as InSar) for specific applications
  - ✓ **Open** – Open Platform, Open Data, Open Computing, Open Source
  - ✓ **Federated** – Leverage existing resources; also through the RSS
- ☒ Widen and deepen scope in ESA Earth Observation Envelope Programme after agreement with ESA Member States



Guiding principles for a federated EO Ground Segment are rather obvious: Simple, Flexible, Portable, and Scalable

Federation addressed along several lines

- Pilot projects to demonstrate use of existing European GS facilities, based on Big Data paradigm
- Federated user management function – evolution of ESA EO Single Sign-On
- Data integrity and certification in a federated context
- Federation among thematically similar capabilities in the TEPs context

- Big Data exploitation in the EO context still pose issues for R&D
- ESA has obtained conclusive results from various initiatives, and continue activities addressing new, complementary operations concepts in GS and ICT:
  - Federated approach to GS and ICT capabilities
  - Managed services in support of Big Data exploitation
  - Thematic Exploitation Platforms
  - ICT provisioning, IaaS, & cloud computing
  - IPR, licensing, payment and funding models
- ESA R&D experience is continuously fed back into the Copernicus Space Segment Ground Segment (CSC GS) architecture, operations concept and evolution

ESA's 2<sup>nd</sup> " BIG DATA FROM SPACE " event foreseen in early 2015

- Focused on Earth observation space/ground and in-situ segments, applications and general Big data exploitation

ESA-wide conference on Research, Technology and Innovation associated to large amounts of \*space\* data

- Addressing all space domains, e.g. technology, planetary, space science, human spaceflight
- Schedule on 12-14 November in ESRIN - Frascati

**Thank you for  
your attention**