Helix Nebula, the Science Cloud

A public-private partnership building a
multidisciplinary cloud platform
for data intensive science

EUDAT 2d Conference, 29 October 2013

Maryline Lengert, ESA

Helix Nebula, the Science Cloud initiative is a **Partnership** between big science and big business to implement the **Strategic Plan for a Scientific Cloud Computing Infrastructure in Europe**

Strategic Plan for a Scientific Cloud Computing infrastructure for Europe

CERN-OPEN-2011-036
08/08/2011

1.1

8th August 2011

1. Set up **a sustainable multi-tenant cloud computing infrastructure** in Europe, initially based on the needs of the European Research Area & Space agencies

2. Identify and adopt **policies** for trust, security and privacy on a European-level

3. Create a light-weight **governance** structure involving all stakeholders

4. Define a short and medium term **funding** scheme

Contacts

Dr. Maryline Lengert
ESA - European Space Agency
Senior Advisor
Maryline.Lengert@esa.int
Tel +39 06 941 80430

Dr. Bob Jones
CERN – European Organization for Nuclear Research
IT department
Bob.Jones@cern.ch
Tel. +41 22 767 14 82

Synergy between HN and EUDAT

EUDAT Vision: support a collaborative Data Infrastructure (cross domain for research)

HN Vision: Provide Cloud Computing capacity and services that elastically meet big science's growing demand in a cross domain environment.

In order to become 'Information as a Service' ecosystem, we need:

- Data e-infrastructure providing sustainable platform of technology, tools, and services
- Platform for shared services that make it possible for data-intensive research to span all the scientific disciplines
- Wide spread access to data and long term-preservation of data for use and re-use

Building on European e-Infrastructures

- Existing European e-infrastructure *long-term* **projects**
  - GEANT, EGI, PRACE, EUDAT

- Many "pathfinder" initiatives have prototyped aspects of what will be needed in the future
  - Includes much of the work in the existing e-Infrastructure projects
  - Thematic projects / infrastructures such as BioMedBridges/ CRISP/ DASISH/ ENVRI, EMSO as well as CORIOLIS / GEOWOW / DORIS and many others

- Future service infrastructure and tools must be fully based on **open standards**, **open software**, and enable **open data access**

Helix Nebula, the Science Cloud

- It is **NOT** a project!

- It doesn't have a paying customer: it is meeting **strategic interest** of each participant preparing the future

- It has been set up by "Extreme Labs", IT & T-Com industry, SMEs (application & computing) to create something new by **bringing together a fragmented landscape** and building on strength from **sustainable diversity**

Strategic Objective

- The European Research Area shall drive the development and implementation of a **secure and globally recognised European Cloud Computing Infrastructure**, initially targeting science users. This infrastructure will become **THE** platform for Europe, under public governance, ensuring **open standard** and **interoperability** and adhering to **European policies, norms and requirements**.

To make this happen, we created Helix Nebula, the Science Cloud Partnership

Partnership – Why?

- The **scale and complexity** of services needed to satisfy the foreseen needs of Europe's IT-intense scientific research & space organisations are beyond what can be provided by **any single company** and will require the collaboration of a variety of service providers and SMEs.

- By partnering with Science and Space Organizations, it shall cover most, if not all, of the needs of a public Cloud infrastructure.

Partantership – Who & How?

IT Providers
-commit resources
-share investments
-agree on standard
-interoperability

Scientific and Space Organizations
-commit res

EC Projects
-Standard/Open source
-Contributing / using HN

Users and Providers
- Act as catalyser

-Policy & Strategy
-through targeted calls under FP7

Partnership has grown from 20 to 38 members in less than 2 years!

Where do we want to go
and
why do we want to go there?

Long Term Goal

- To create a multi-tenant '**Open Market Place for Science**', where data, scientists, funding bodies, SMEs and downstream industry are meeting to work along common interests.
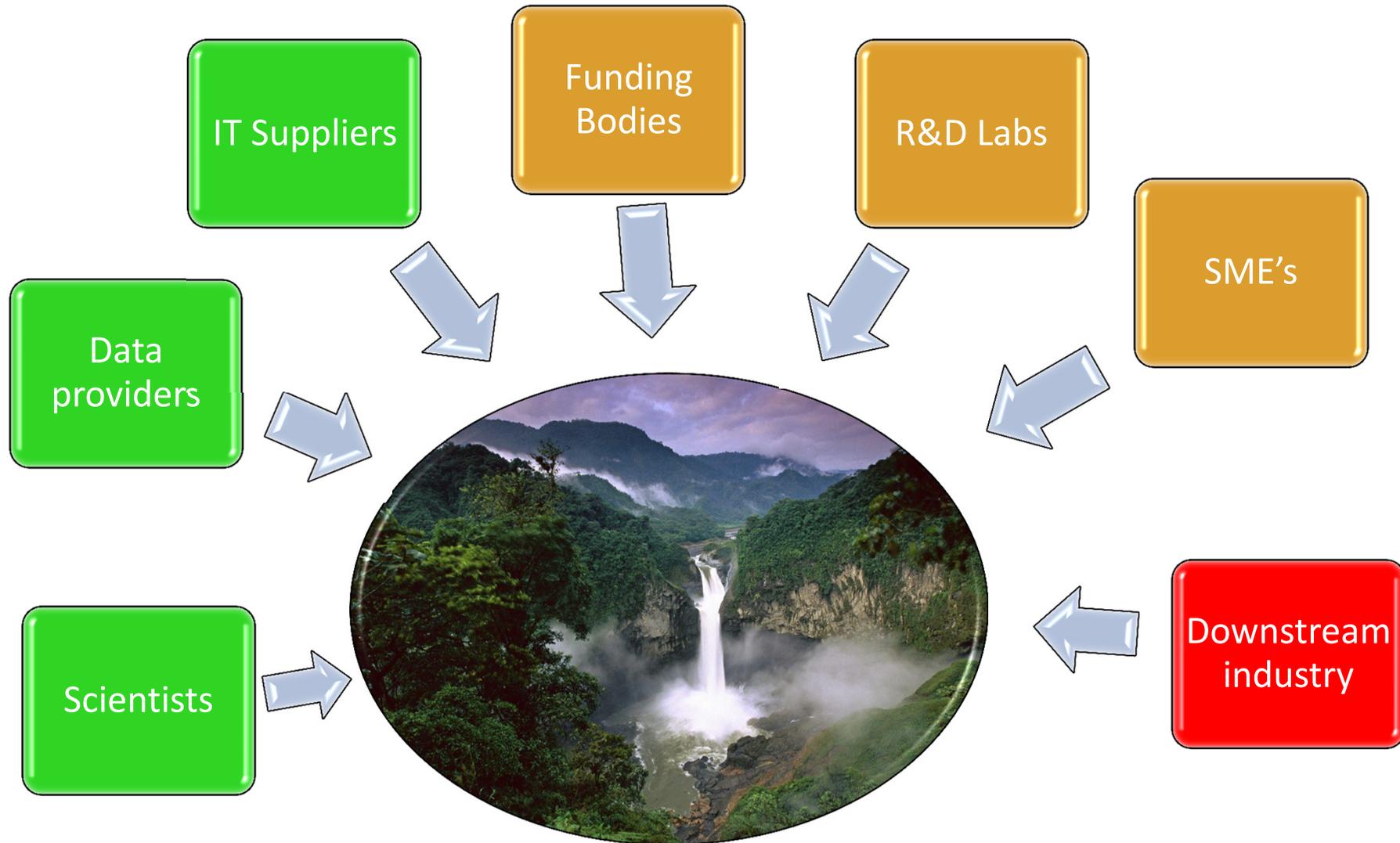
**This ecosystem should implement
many-to-many relationships, quickly being
established, to transform
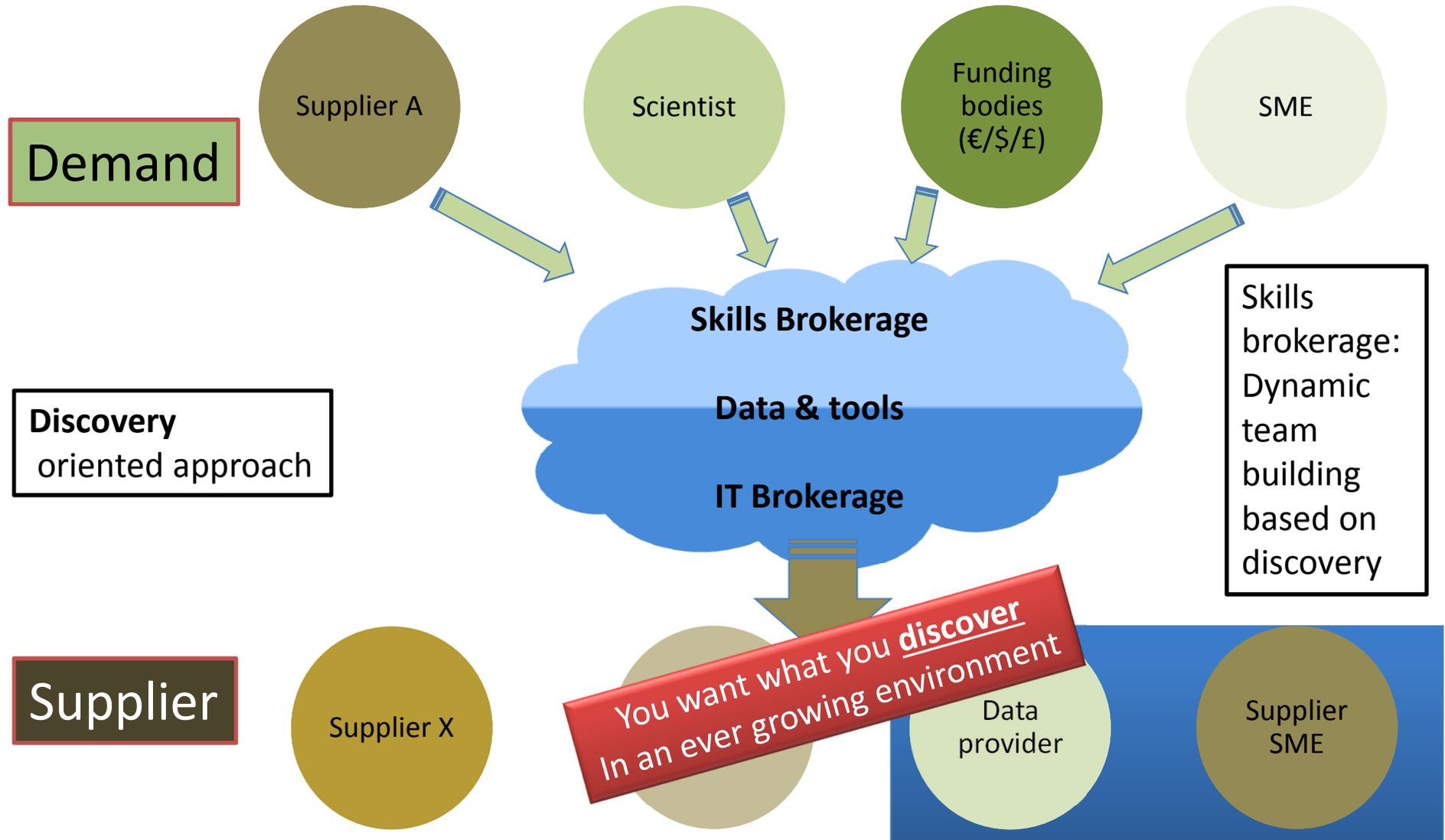data into valuable information**

Operating as an Ecosystem

An **ecosystem** is a community of living organisms (plants, animals and microbes) in conjunction with the nonliving components of their environment (things like air, water and mineral soil), interacting as a system.
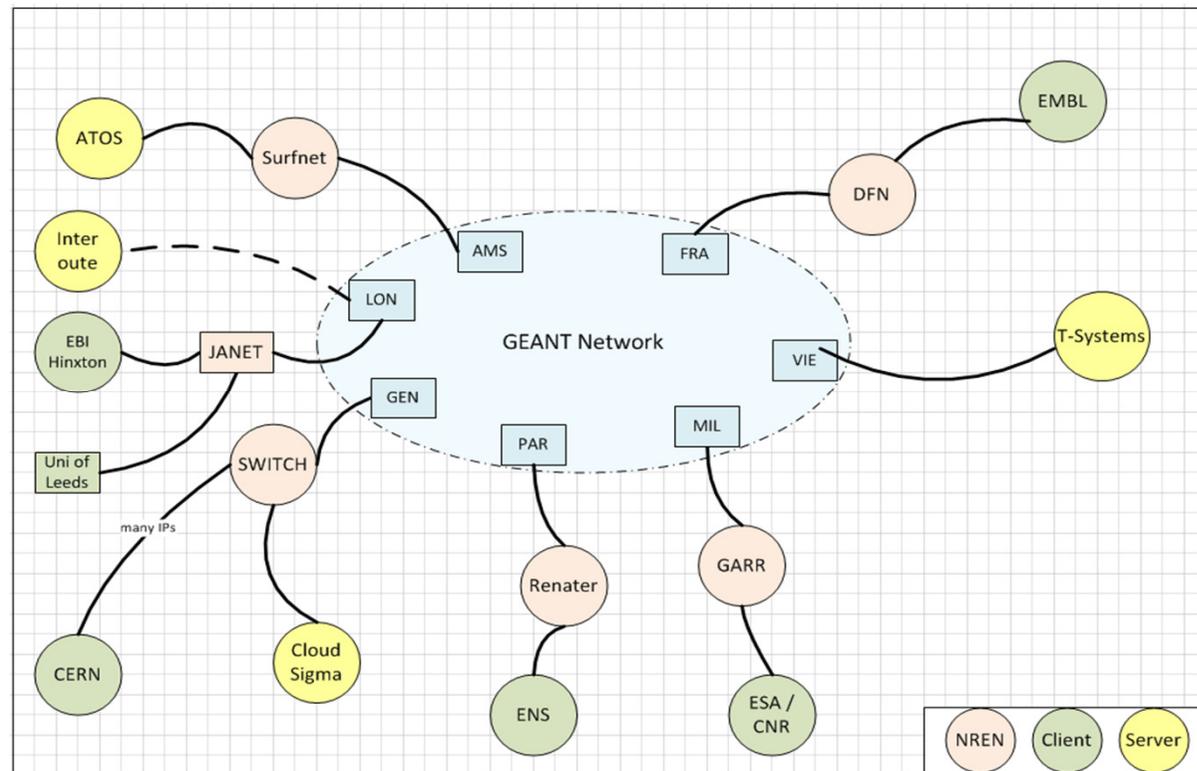
Contributors to the Ecosystem

IT Suppliers

Funding Bodies

R&D Labs

SME's

Data providers

Scientists

Downstream industry

Final Ecosystem: Discovery driven

Demand

Supplier A

Scientist

Funding bodies (€/$/£)

SME

Skills Brokerage

Data & tools

IT Brokerage

Discovery oriented approach

Skills brokerage: Dynamic team building based on discovery

Supplier

Supplier X

You want what you **discover** In an ever growing environment

Data provider

Supplier SME

Building the hybrid cloud
Connecting commercial providers to the GÉANT/NRENs

DANTE offering free IP connectivity in GÉANT for research traffic during the pilot phase

NRENs have different commercial agreements (usually they apply a fee)

Helix Nebula and Public Infrastructures proposed hybrid model

Academic

Big Science   Long tail

Other market sectors

Government   Manufacturing   Oil & gas, etc.

HN as supplier to research

EGI

Fed Cloud Task Force

Blue Box

Helix Nebula

Atos   Cloud Sigma   T–Systems

Public funded

Commercial

Network Commercial/GEANT
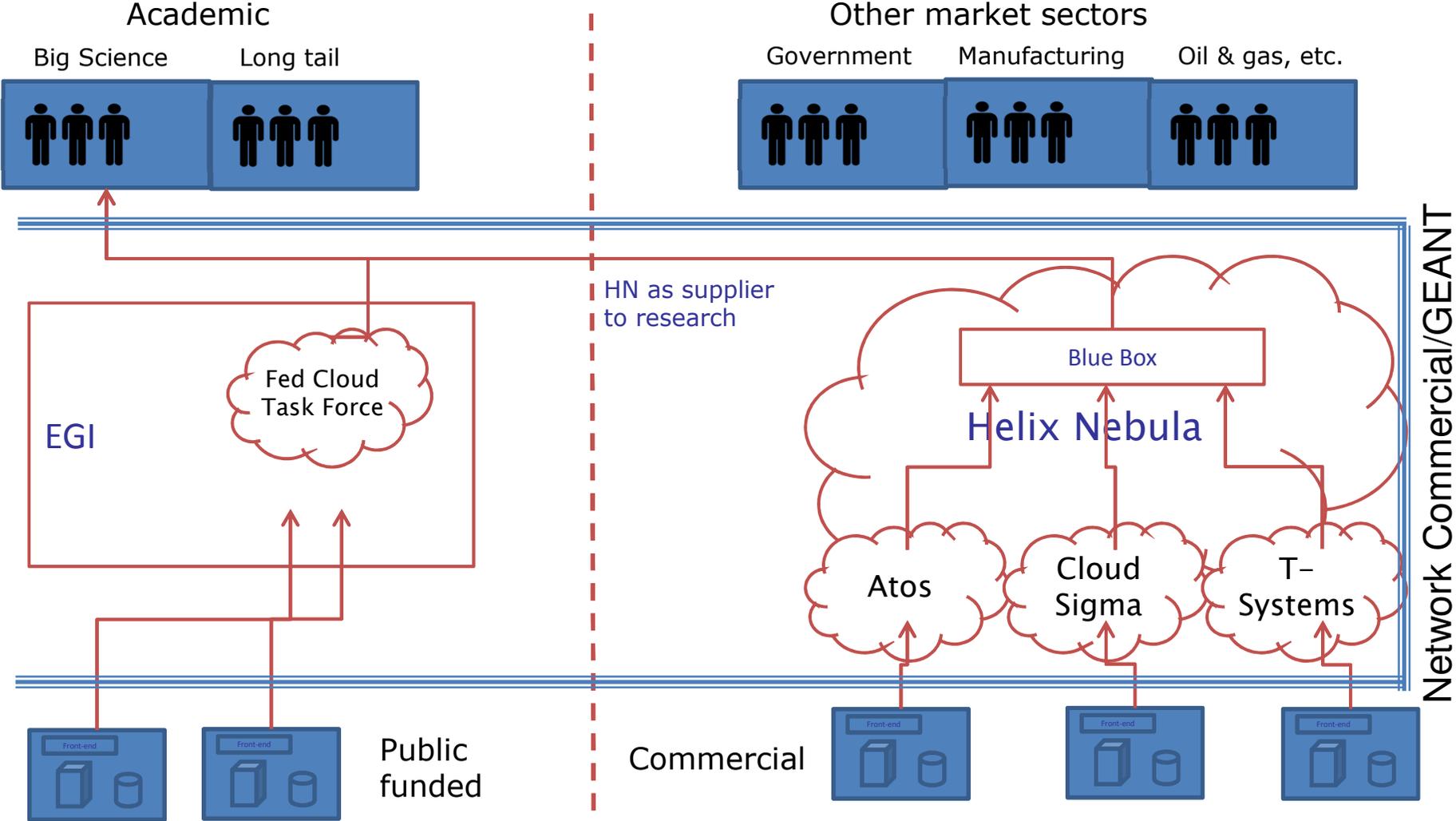
Expected Benefits

- We build **scale** to allow you and us to handle **elasticity**

- **Risk sharing** between partners

- Compliance with **EC rules and policy** (privacy, security, ethics)

- **Business opportunities** for SMEs

- **Breakthrough in Science** thanks to cross-domain fertilization (data, tools, funding, capacity, peers)

Validating our approach

Timeline

2011

- Endorse the Common **Strategy**
- Agree on the **Partnership**
- Select **flagships** use cases
- Define **governance** model

2012-2013

- **Pilot** Phase
- **Deploy** flagships,
- **Analysis** of functionality, performance & financial model

2014 ...

Start Operations Towards an **open market for Science**

Initial Flagship Use Cases

ATLAS High Energy Physics Cloud Use
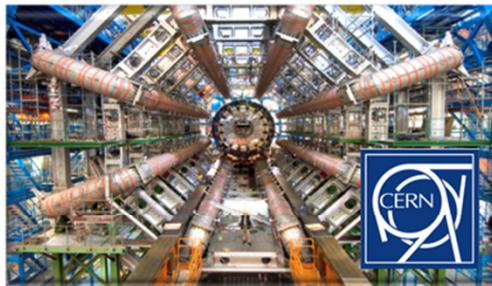
To support the computing capacity needs for the ATLAS experiment

Genomic Assembly in the Cloud
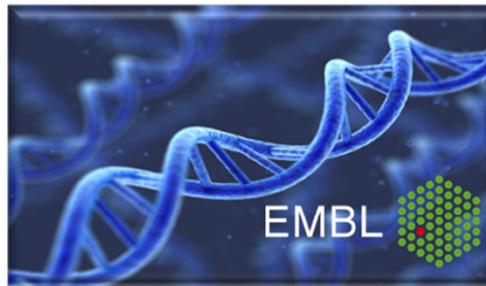
A new service to simplify large scale genome analysis; for a deeper insight into evolution and biodiversity

SuperSites Exploitation Platform

To create an Earth Observation platform, focusing on earthquake and volcano research

- Scientific challenges with societal impact
- Sponsored by user organisations
- Stretch what is possible with the cloud today

ESA Flagship
SuperSite Exploitation Platform

What is a Super Site?

Data sharing (**all** data space & in-situ) accessible at no cost through a one-stop shopping e-infrastructure (containing also tools and peers).

Cross-domain: The supersites will lead to a new, more versatile generation of scientists that capitalize on information from multiple techniques.

International collaboration with regional responsibilities:
The backbone of the Supersites initiative is global collaboration of scientists and data providers. National organizations should coordinate efforts for establishing regional integrated research infrastructure in charge of long-term integration plans such as European Plate Observing System (EPOS) for Europe
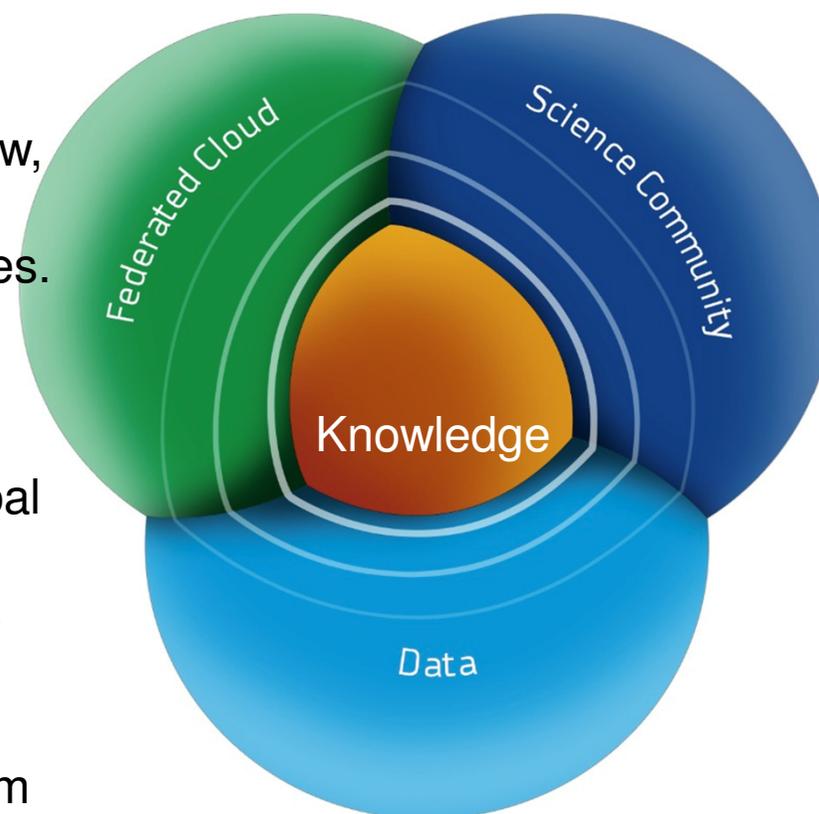
The Geohazard Supersites

Support GEO to better understand the geophysical processes causing **Geohazards** (earthquakes and volcanoes)

- Global partnership of scientists, satellite and in-situ data providers (**multi-sensor InSAR, seismic, GPS - complete data sets**)

⇒ Brings together community & relevant data

- Support national authorities and policy makers in risk assessment and mitigation strategies for Geohazards

Group on Earth Observations

Intergovernmental Organization with 88 country members and

67 participating organizations (as of March 2013)

Construct by 2015: **G**lobal **E**arth **O**bservation **S**ystem of **S**ystems (**GEOSS**)

U.S. Department of State, Washington DC
July 31, 2003

GEO data sharing principles approved in 2010 Plenary: open & free data access for science

GeoHazard Supersites: Infrastructure concept

EO Science users will use the Supersites their first option to store and access data, for data processing and analysis because that's where they find all the data, tools and peers.
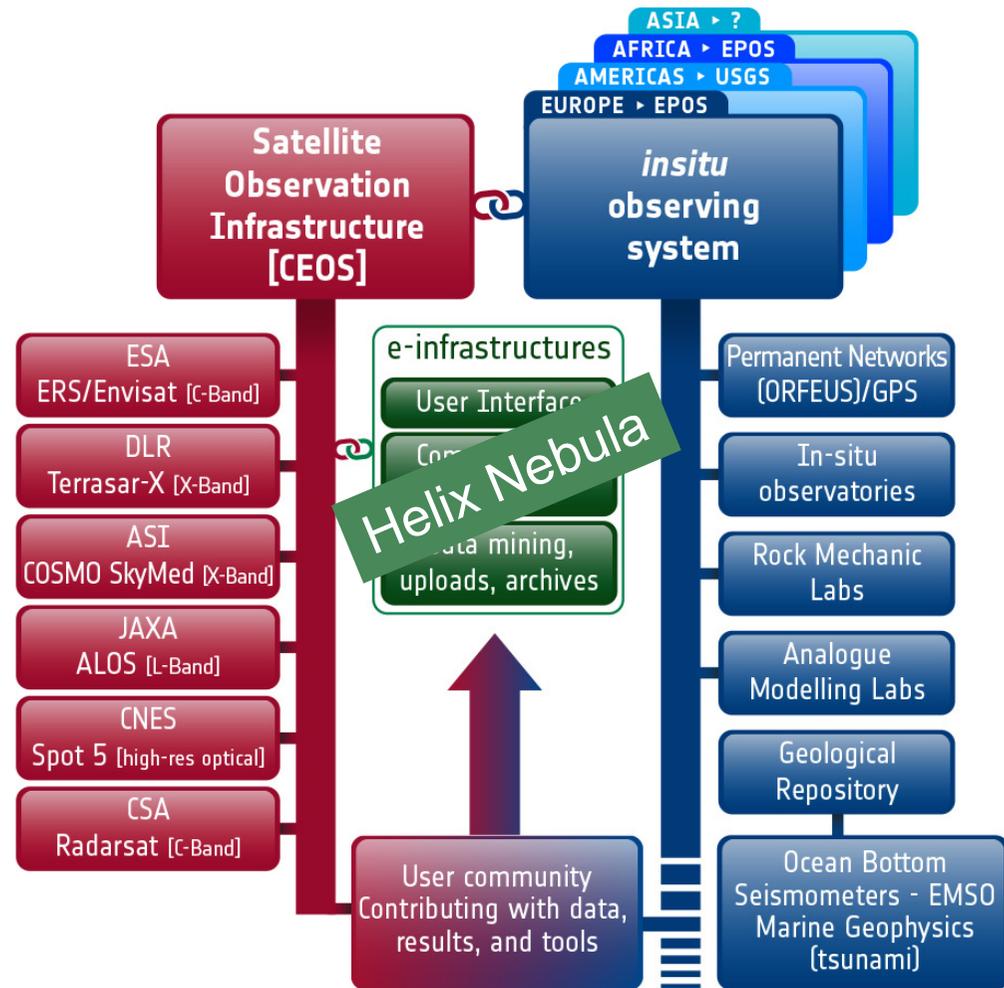
SSEP Expected Results

1. Science: better scientific understanding of geohazards with the aim of providing sound information about the risks and the potential mitigation measures

2. Data sharing: information extracted from different sources (satellite & in-situ) will open a wide range of new approaches: Cross-domain research

3. Building Communities : In return to SSEP access, scientists will be asked either:

   - to provide results into the information repository on the Science Cloud,

   - to provide their data processing open source code,

   - to provide application tools ("Apps")

Geo Hazard : Japan earthquake

EARTHQUAKES OF THE JAPAN TRENCH
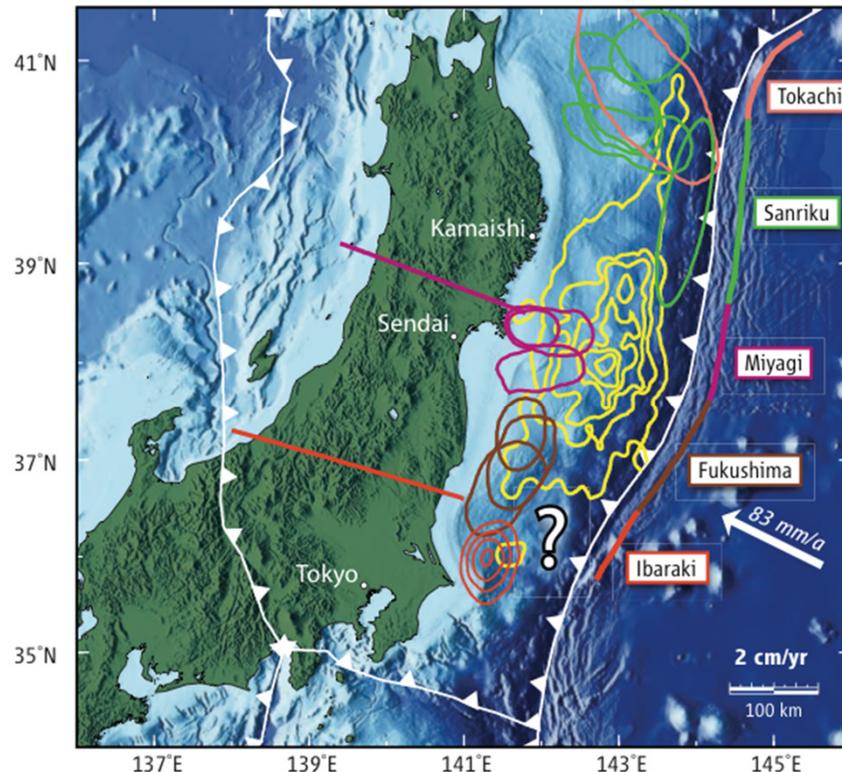
A game of ring toss. March's huge quake (yellow contours) and past smaller quakes (colored loops) have left a patch of threatening fault (question mark).

- Tohoku-oki: unprecedented >50 m slip in places (Simons et al., Science 2011, NASA-funded study).

- Will another magnitude 9 occur further south?

- It is unknown whether this fault segment has been accumulating slip.

- Need all InSAR, GPS, Seismic, Petrology, Geochemistry , ... !

(-2004 magnitude 9.2 Sumatra earthquake was followed by magnitude 8.7 half-a-year later)

→ **The Science Cloud with its "unlimited" resources on data, processing capacity and tools will allow cross-domain science and ease data sharing. The easy usage of this infrastructure will pull "intelligence" to apprehend the challenges.**

Earth Observation Application Platform exploiting 20 years of satellite data

- EO Application Platform
  - OpenNebula
  - Data Catalogue and Access
  - Map-Reduce computing model
  - Software repository
  - Utilities for sw development and testing
- Cloudification of application
  - CNR / IREA (Italian Research Council in Naples) developed an application (SBAS) measuring the vertical movement of ground in sub cm from space.
  - **SBAS** targets
    - Time series over **20 years** with ESA archive
    - Points of Interest are at **world scale**
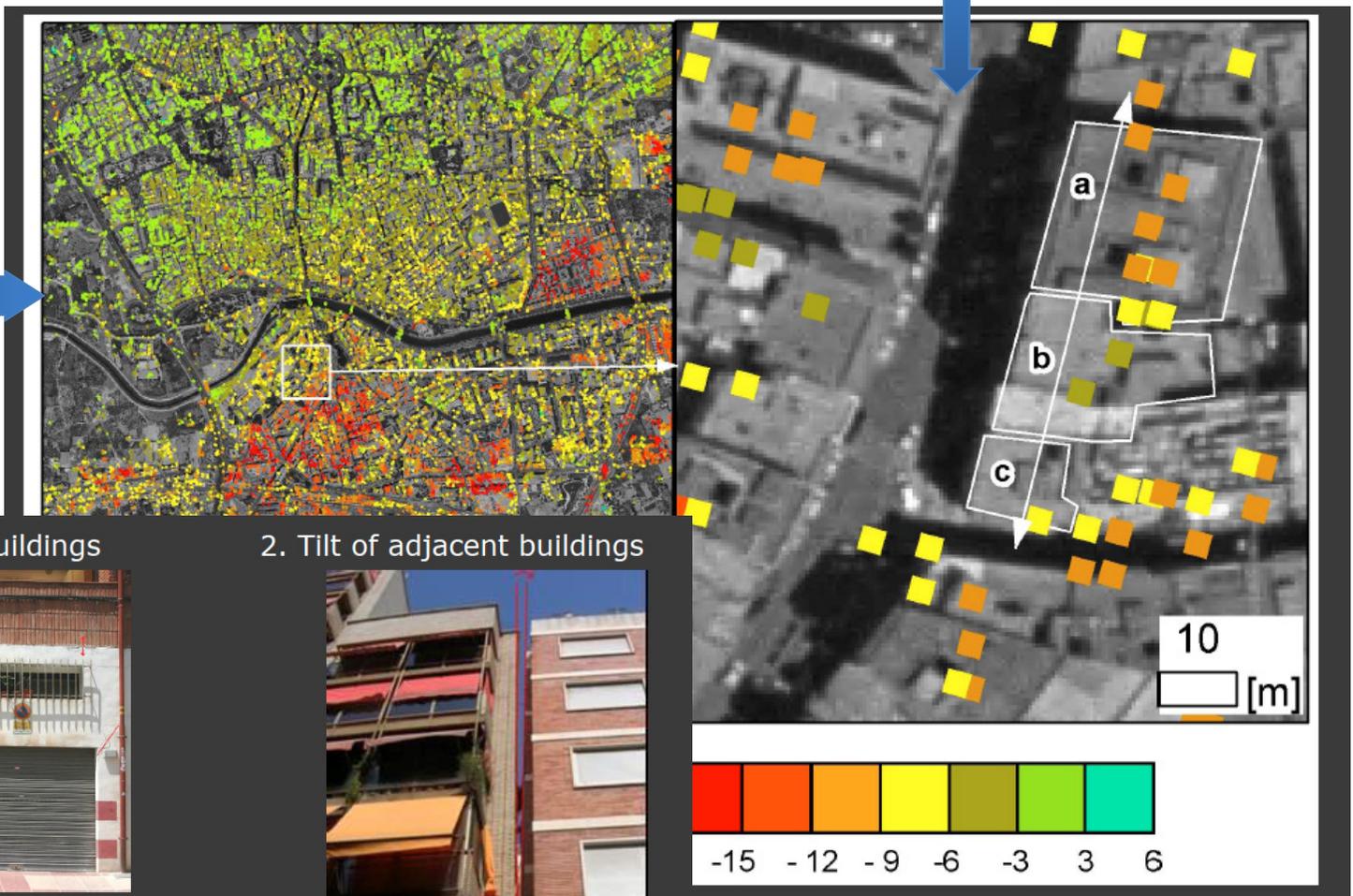   - **TBytes** of data to process
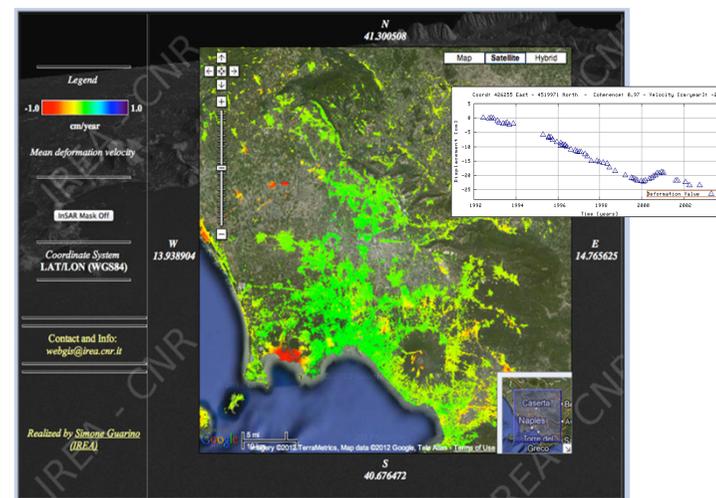
A crack in a house…

His Res commercial
Data policy

ESA open and free
Data policy

In situ data from
various providers

1. Tilt of individual buildings

2. Tilt of adjacent buildings

3. Negative skin friction on foundation piles

a

b

c

10
[m]

-15 -12 -9 -6 -3 3 6

Courtesy from DORIS

Grant Agreement #: 242212

Data for geo-hazard leading to "Information as a Service"

- **Satellite data: structure, catalogued**
  - ESA data with medium res with open and free data policy (20 years of data in the archive)
  - Commercial data TerraSAR-X, COSMOSkyMed, Radarsat
  - High res optical data (IKONOS, GEOEYE, Quickbird, SPOT-5, FORMOSAT-2, KOMSAT-2, Pleiades …)

- **In-situ data:**
  - Structured data:
    - of Geological Surveys (e.g. EPOS) (Inclinometer, GPS, seismometer,..)
    - from scientific & commercial measurements and national authorities
    - Aerial fotos
    - Digital Elevation Models
    - "Did you feel it?"   (e.g. App form USGS for Earthquake)
    - Ground deformation maps and time series,
    - Displacement & damage maps

  - Unstructured data
    - e.g. footage with meta data from Tsunami event

Satellites data

32

Courtesy from DORIS

Maryline Lengert, ESA
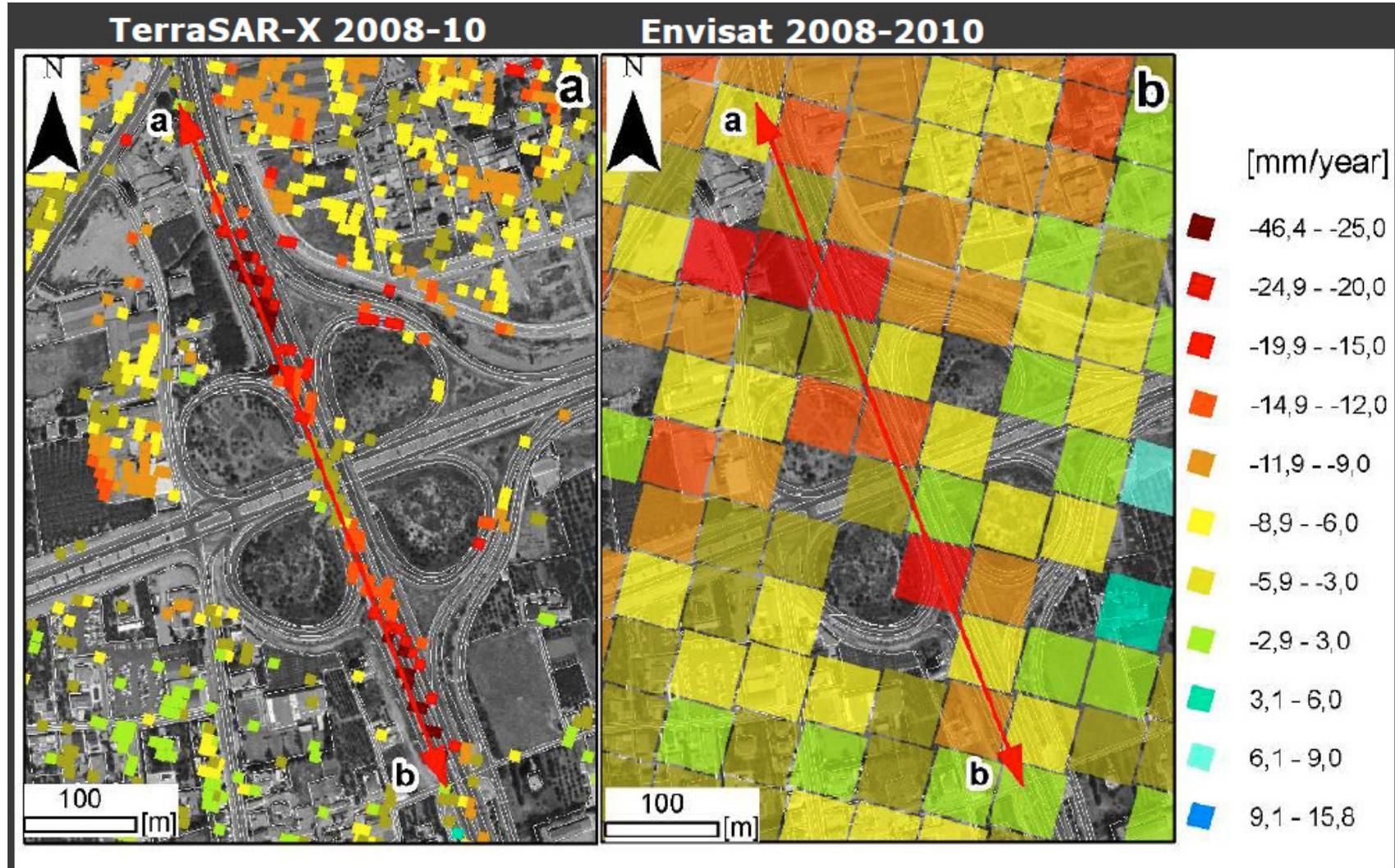
A European cloud computing partnership:
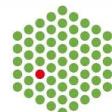# big science teams up with big business

Strategic Plan

▶ Establish multi-tenant, multi-provider cloud infrastructure

▶ Identify and adopt policies for trust, security and privacy

▶ Create governance structure

▶ Define funding schemes

To support the computing capacity needs for the ATLAS experiment

Setting up a new service to simplify analysis of large genomes, for a deeper insight into evolution and biodiversity

To create an Earth Observation platform, focusing on earthquake and volcano research

Adopters

http://www.helix-nebula.eu
contact@helix-nebula.eu

@HelixNebulaSC
HelixNebula.TheScienceCloud