EPOS PP

European Plate Observing System | FP7 Preparatory Phase Project

Seismology and EUDAT

(…an EPOS community case)

Alberto Michelini with many others

INGV, Rome, Italy

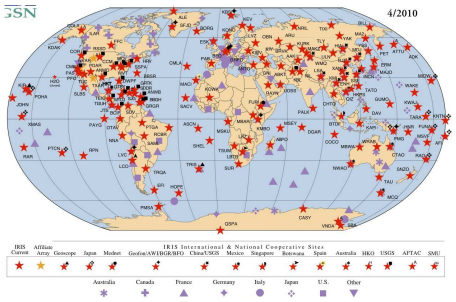
www.epos-eu.org

EUROPEANPLATEOBSERVINGSYSTEM

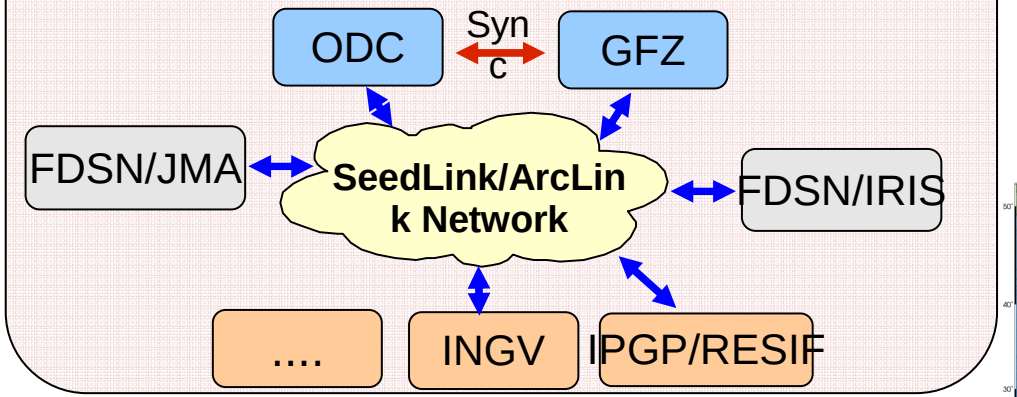
What is Seismology ?

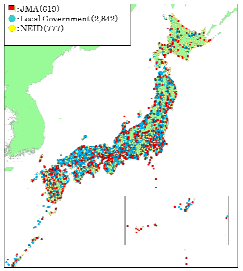
- Scientific field that studies earthquakes *(standard definition)*
- Scientific field that studies seismic transients within the Earth using seismogram recordings primarily *(more up-to-date definition)*
- Seismology is likely the most important science for investigating the interior of the Earth because of the intrisic resolving power provided by the seismic data

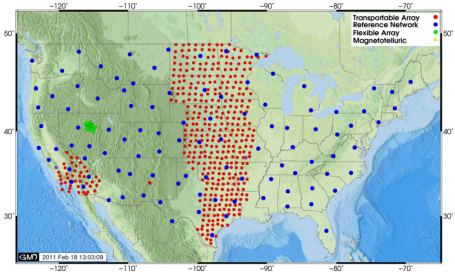
*Seismology is the most advanced community within **EPOS***
(European Plate Observing System)



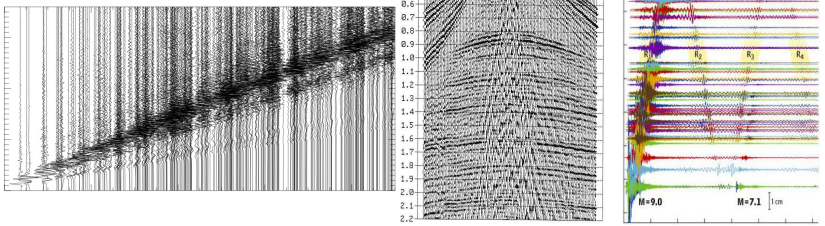
EUROPEAN INTEGRATED DATA INFRASTRUCTURE (EIDA)







SEISMIC RECORDS



Data and Modeling Intensive Research

Visualization,
Data analysis /parallel data mining
Simulation, inversion,
HR imaging

PRACE

HPC/GkiD

Infrastructures

EGI

Courtesy of Jean-Pierre Vilotte, VERCE

Data Intensive applications

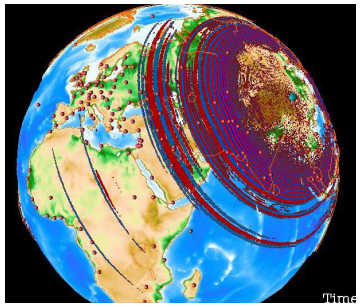
Earthquake and seismology community is facing a fundamental paradigm shift: from data driven to data intensive research:

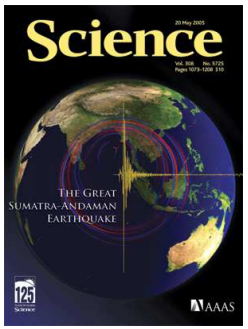
Large volume data analysis: extracting information from space and time correlations in dense array observations,
Data and computing intensive simulation/inversion: 3D wave form information using adjoint methods, stochastic strong motion simulation,
Orchestrated workflows across service components.

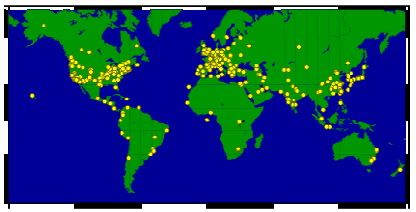
Seamless access to large volumes of multi-sets data across the Grid and HPC components

Industrial and societal applications: natural hazards, climate changes and energy resources and national security.

Large earthquake source radiation: Sichuan (Mw 7.9, 2009, China); Sumatra-Andaman (Mw 9.2, 2004, Indonesia)

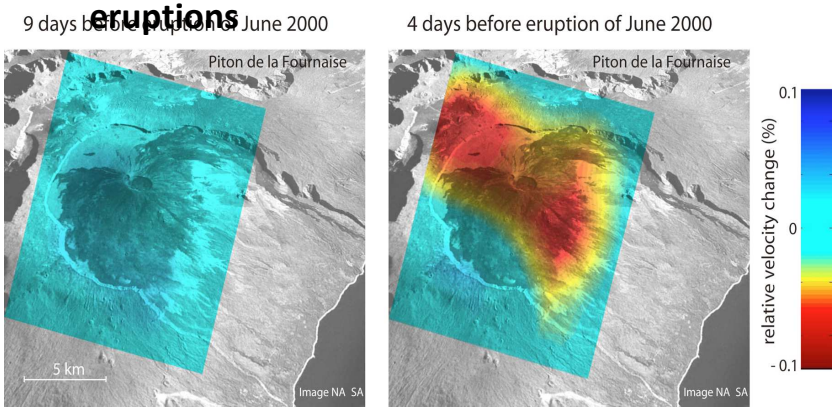




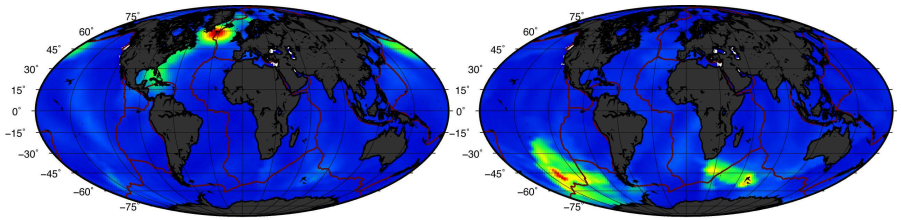


Research groups using SPECFEM3D

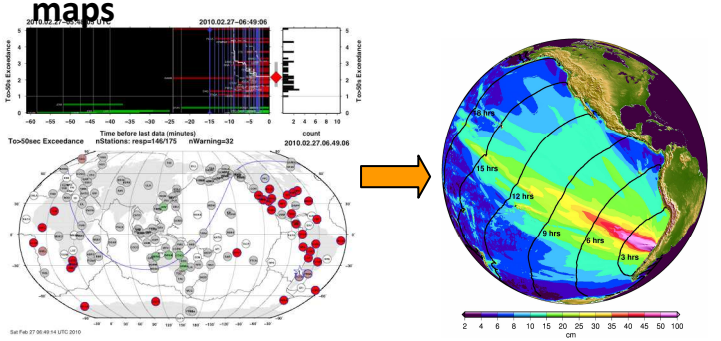
Seismic noise correlations: observing precursors to volcanic eruptions



Studying the coupling between the Solid Earth, the Oceans, and the Atmosphere



Earthquake detection: tsunami impact maps



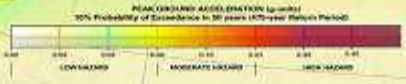
Courtesy of Jean-Pierre Vilotte, VERCE

What is EPOS ?

EPOS is a long-term integration plan that aims to create a single sustainable, permanent and distributed infrastructure that includes:

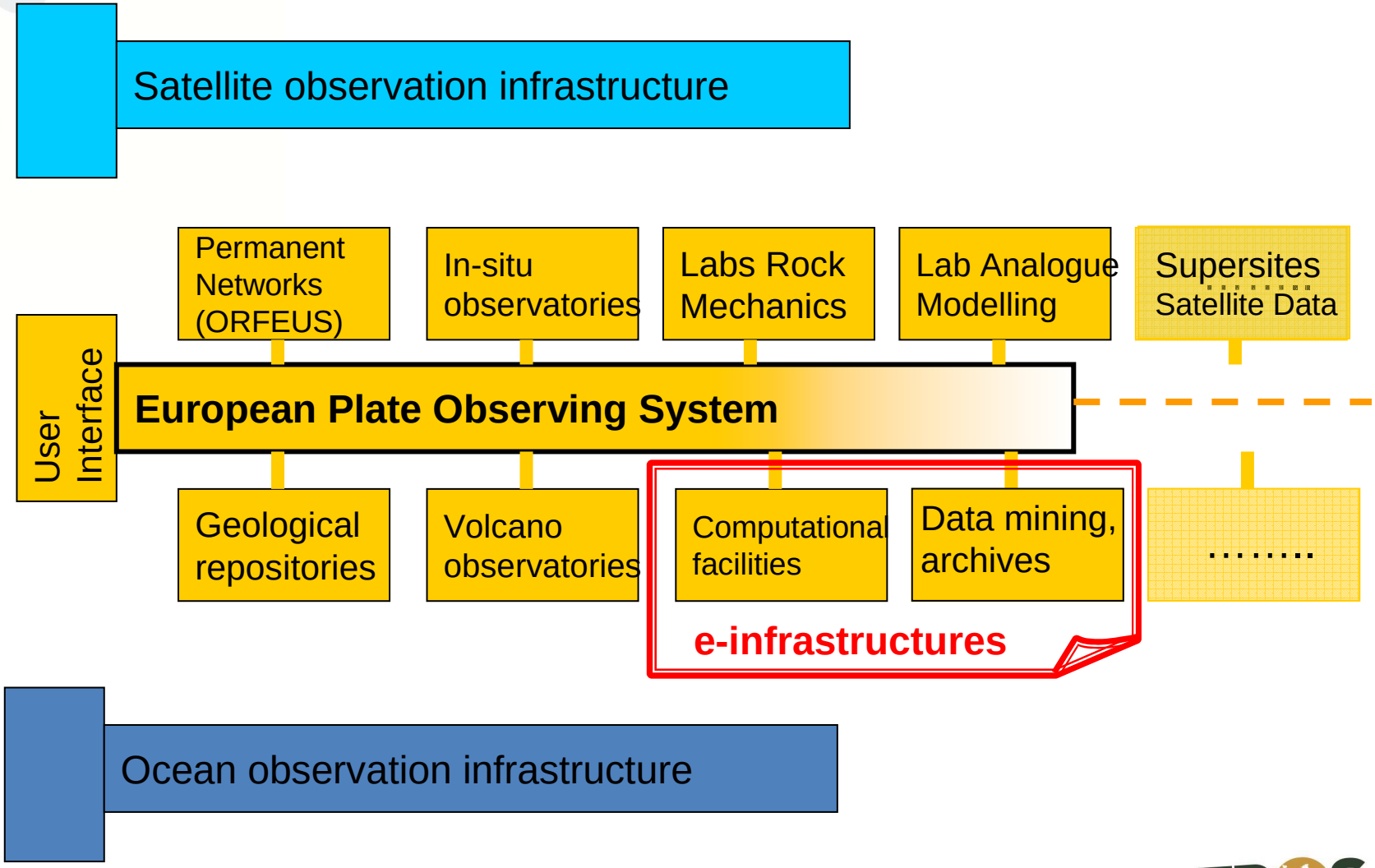
- geophysical monitoring networks
- local observatories (including permanent in-situ and volcano observatories)
- experimental & analogue laboratories in Europe

EPOS will give open access to geophysical and geological data and modelling tools, enabling a step change in multidisciplinary scientific research into different areas



seismic hazard map

EPOS infrastructure concept

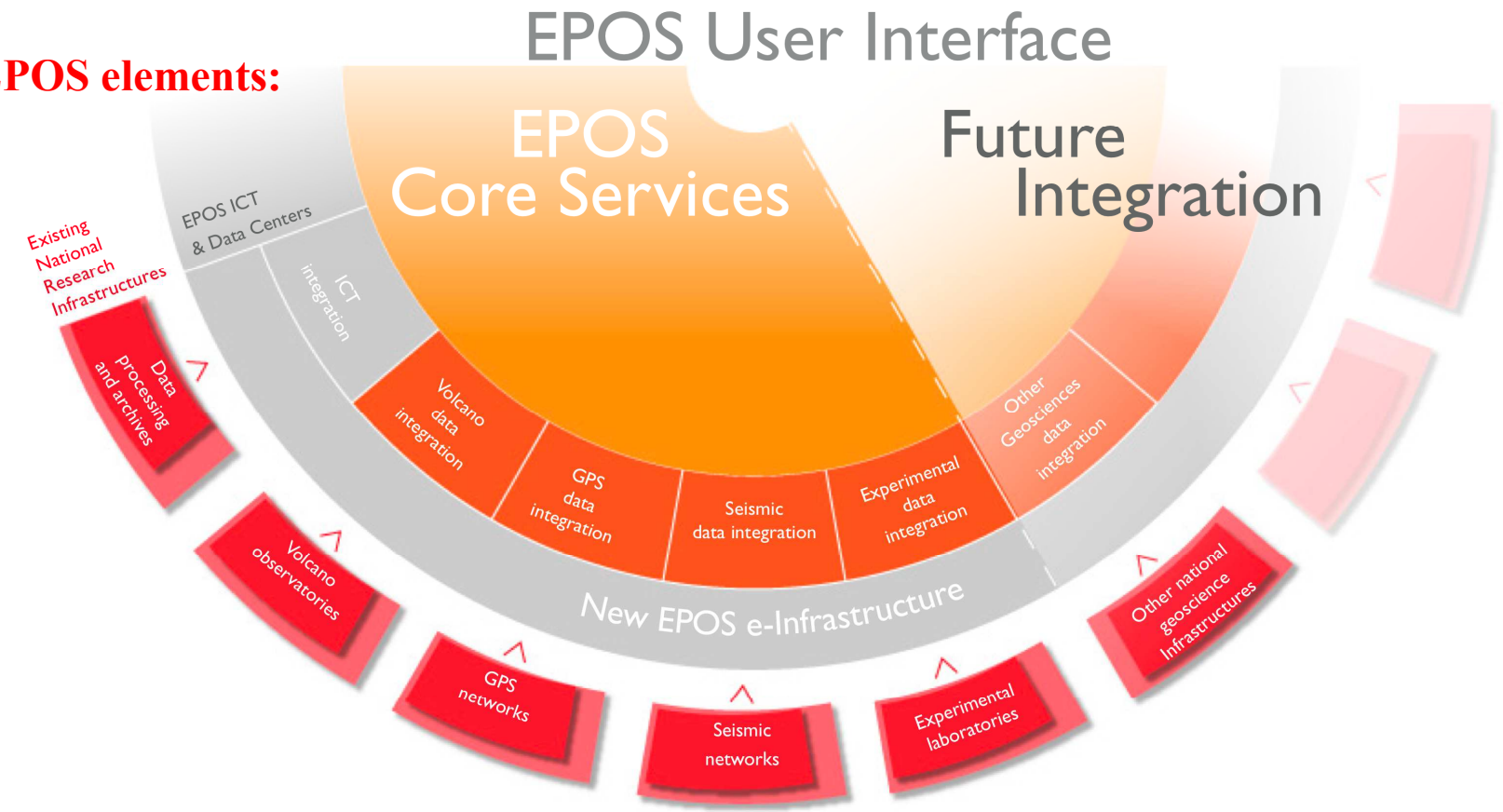


EPOS
EUROPEANPLATEOBSERVINGSYSTEM

Research Infrastructure and E-science for Data and Observatories on
Earthquakes, Volcanoes, Surface Dynamics And Tectonics

How?

The EPOS elements:



- The existing national research infrastructures are integrated into the EPOS Data Centres, which represent community specific services for data archiving and mining having their own computational resources.
- Community specific data centres are further integrated by the EPOS Core Services, representing the infrastructure layer consisting of common data services.
- EPOS data service infrastructure will be designed and established during the PP to serve multiple communities studying the solid Earth dynamics.

What EPOS/Seismology has already ?

- High quality digital broadband seismic data for the last ~25 years
- Community specific archives worldwide and a federated data archive (EIDA) in Europe
- Community services for data extraction and download

What EPOS/Seismology asks to EUDAT ?

A good, reliable, multipurpose and fast “engine” that sits “under the hood” of the community services

- Data replication/synchronization of large data volumes
- Data staging to allow for successive data analysis
- **Metadata for search and mining**
- **Persistent identifiers**
- AAI

Some fundamental questions

- How do I find the data?
- Are there mechanisms to uniquely identify and share the data?
- Can I select the data I want efficiently minimizing data transfer and shipping?
- Can I compare data quality avoiding to download and process it?
- Which are the characteristics that could be of interest for a broad community and which are in converse the personal needs of the users?
- How do I publish and have the data persist the in the long term?

Metadata for search and mining (1)

- The seismology community is somewhat behind when compared to other communities (probably too self-referential!).
- Fundamental issues such as the **discovery**, **evaluation** and **re-use** of data had not been really addressed until now.
- Seismic waveforms are continuously recorded by sensor disseminated all over the world and then stored and archived in a few main data centers.
- Seismology has recently opened a discussion on these topics and it is working toward global consensus on metadata and a discussion group has been started (http://eidawiki.orfeus-eu.org/)

Metadata for search and mining (2)

- Geospatial nature of seismic data requires metadata that includes elements for identification that are
  - Spatial, temporal extent and instrument parameters
- and possibly additional attributes to cover
  - topic, quality, lineage and usage constraints
- This information can afterwards enable different types of search e.g. by keywords, names, free text etc.
- Several schemas and standards available for this purpose (e.g.: Dublin Core, ISO 19115, CERIF, INSPIRE) [some of them are the result of many years of best practices in specific communities others come from international broad consensus efforts]
  - **Most important issue is have a clear idea of the crucial information that we want to represent**

Persistent identifiers

- Need to re-use the data
  - mechanisms to uniquely identify and share the data
  - Mechanism to gather and (re)select the data

- **The discussion has started and there are some solutions are under investigation**

General consideration from the community side

In order to have a successful metadata model that could enable sharing and common understanding of the data, it is important to share the same vision. In the selection of the main features we have to consider some aspects:

- ✓ computational cost. It has to be relatively easy to extract the parameters directly at the data centers dealing with the current architectures
- ✓ storage and management should be affected the least possible
- ✓ current acquisition and ingestion systems

EUDAT and synergy with other projects