VERCE : Virtual Earthquake and Seismology Research Community in Europe e-science environment

Centre National de la Recherche Scientifique (CNRS-INSU), France

University of Edinburgh (UEDIN), United Kingdom

Royal Netherlands Meteorological Institute (KNMI-ORFEUS), Netherlands

European-Mediterranean Seismological Centre (EMSC), France

Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy

Ludwig-Maximilians-Universität (LMU), Germany

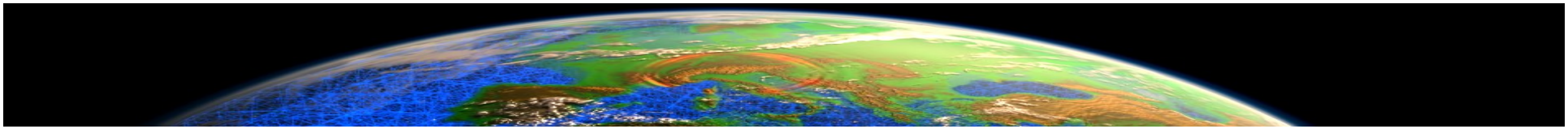
University of Liverpool (ULIV), United Kingdom

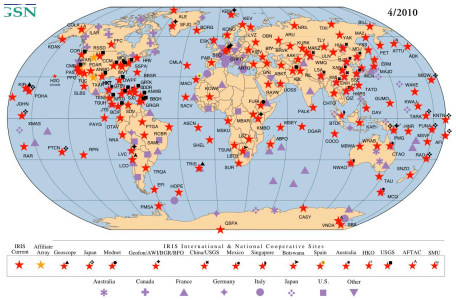
Bayerische Akademie der Wissenschaften (BADW-LRZ), Germany

Fraunhofer-Gesellschaft e.V. (SCAI), Germany

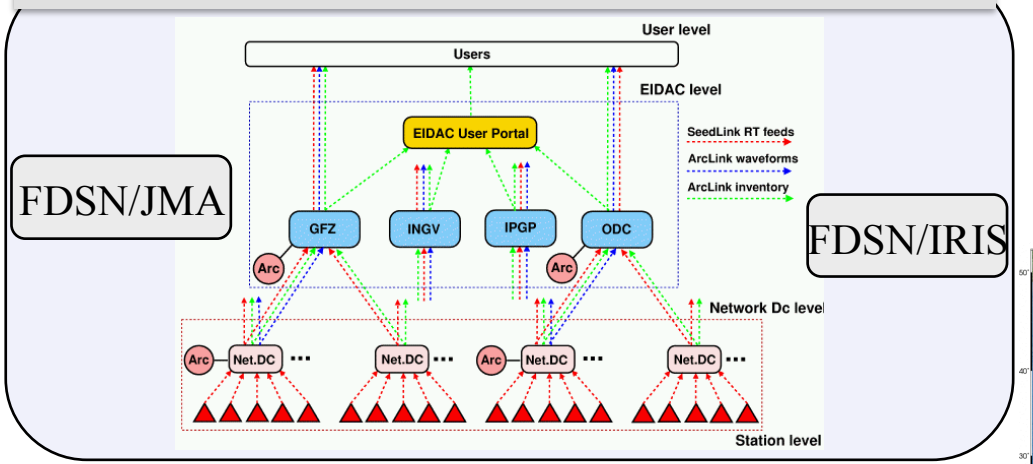
Centro di Calcolo Interuniversitario (CINECA), Italy

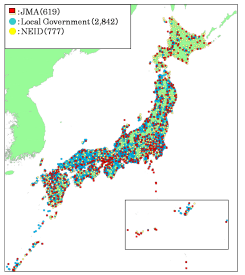
Jean-Pierre Vilotte (CNRS-IPG Paris), Malcolm Atkinson (UEDIN), Torild van Eyck (ORFEUS-KNMI), Anton Frank (BADW-LRZ)

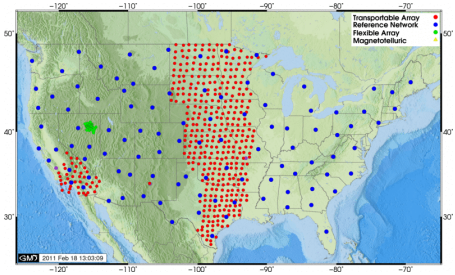




EUROPEAN INTEGRATED DATA ARCHIVES (EIDA)







Data Intensive Research

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

VERCE

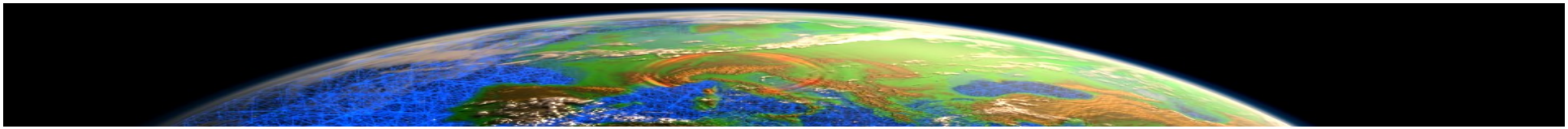
e-Science environment for data intensive research based on an extensive service-oriented architecture

HPC/GRID Infrastructures

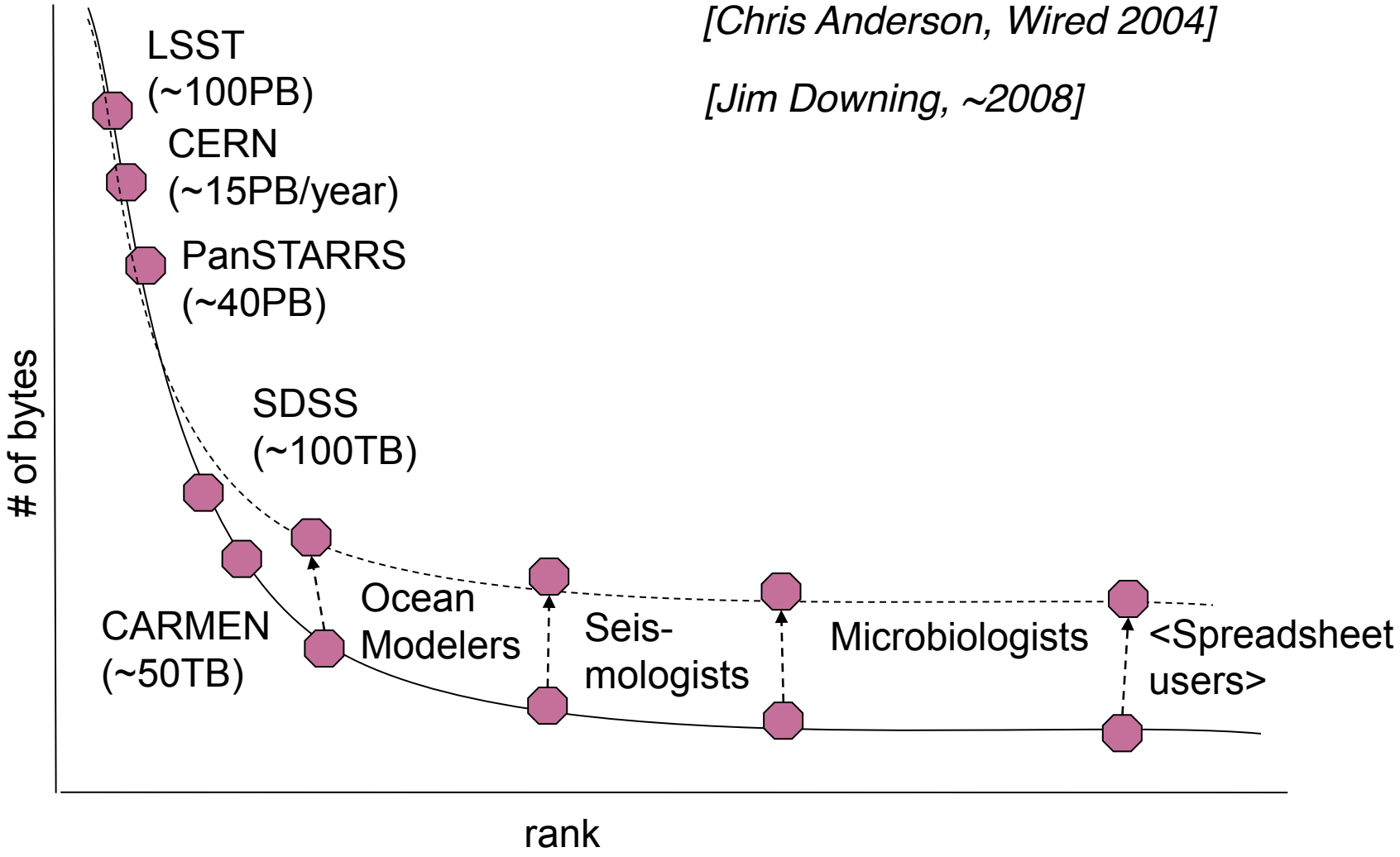
PRACE

EGI

Earth's interior imaging and dynamics: noise correlation, waveform analysis	**Natural hazards: new tools for monitoring earthquakes, volcanoes, and tsunami**	**Interaction of solid Earth with Ocean and Atmosphere: environment, climate changes**

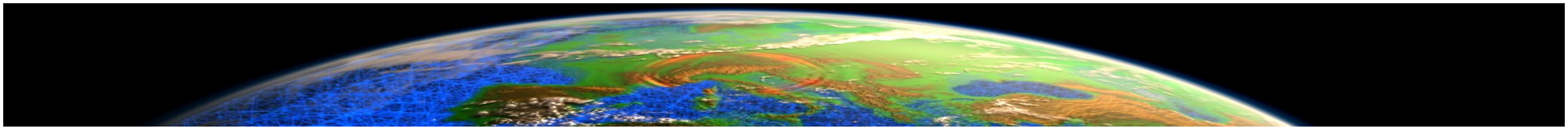


The long "Heavy" tail

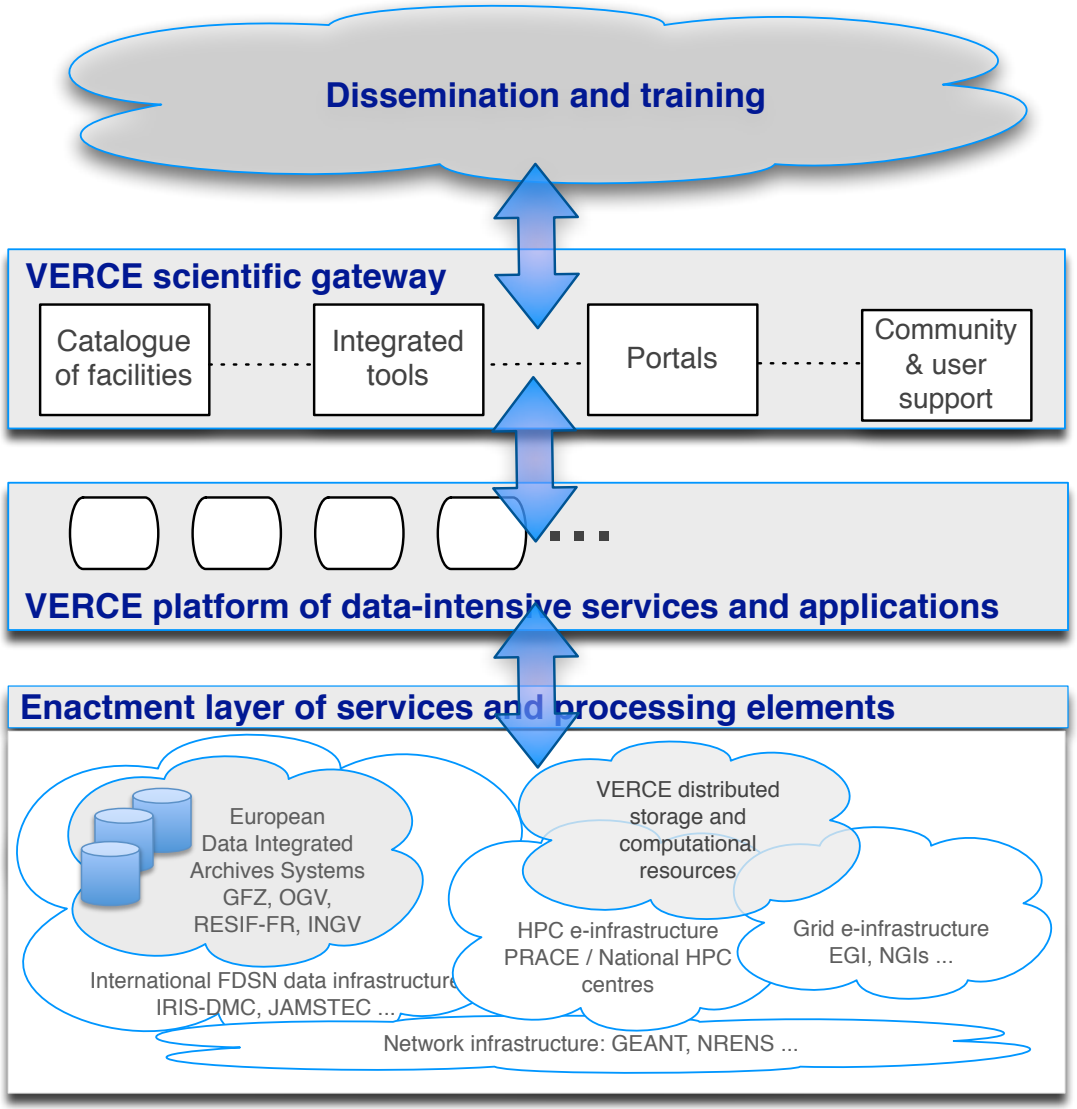


[Chris Anderson, Wired 2004]

[Jim Downing, ~2008]



A service oriented platform



Technology Stack

Web Portal:
Jetspeed, Rapid

Workflow Enactment:
ADMIRE

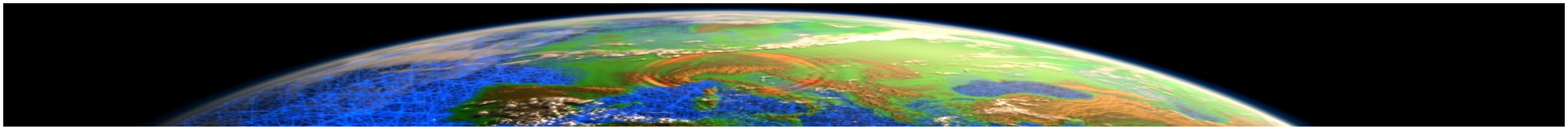
Service & Interoperability:
OGSA-DAI, SAGA, DRMAA

Coupling & Execution:
Kepler, MUSCLE, GridSpace

Data Infra:
Arclink, NetCD, iRODS

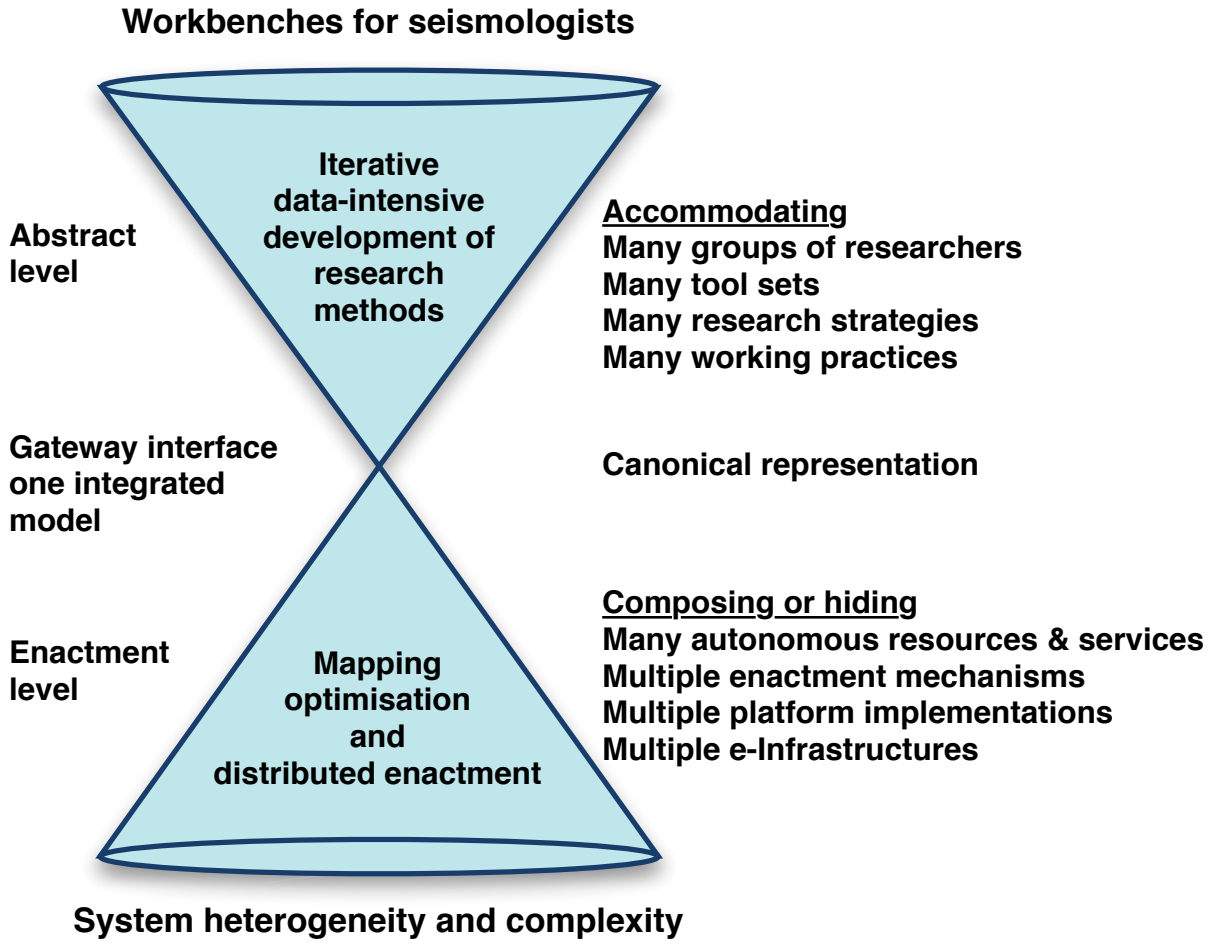
Grid & HPC Infra:
gLite, UMD, UNICORE, OMII-Europe

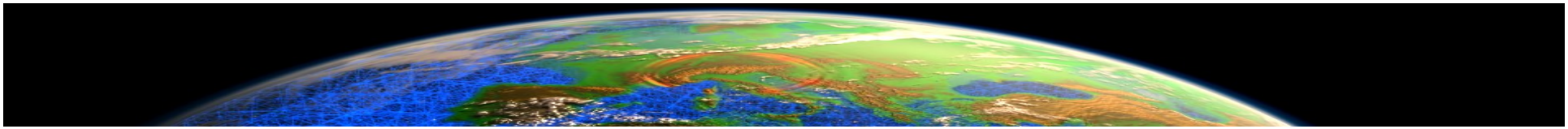
Federated AAI, single sign-on:
Shibboleth, SAML, SLCS, VOMS



A three layer architecture

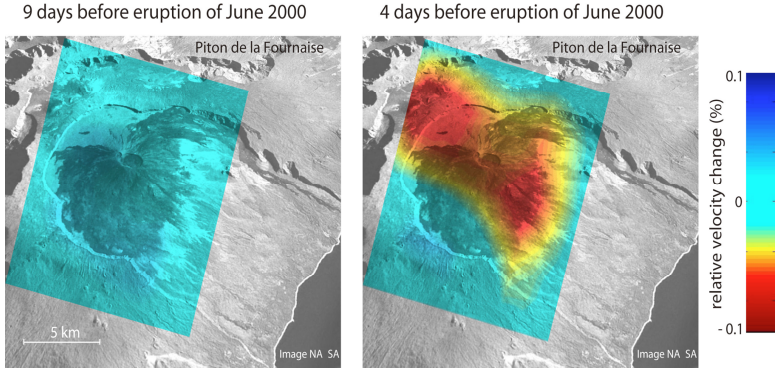
- Separation of concerns
- Resilience toward standards evolution
- Collaborative tools for seismology experts





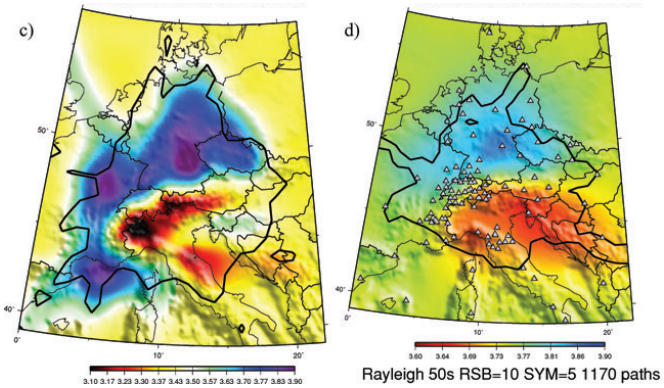
Data Intensive analysis

Seismic noise correlations: observing precursors to volcanic eruptions

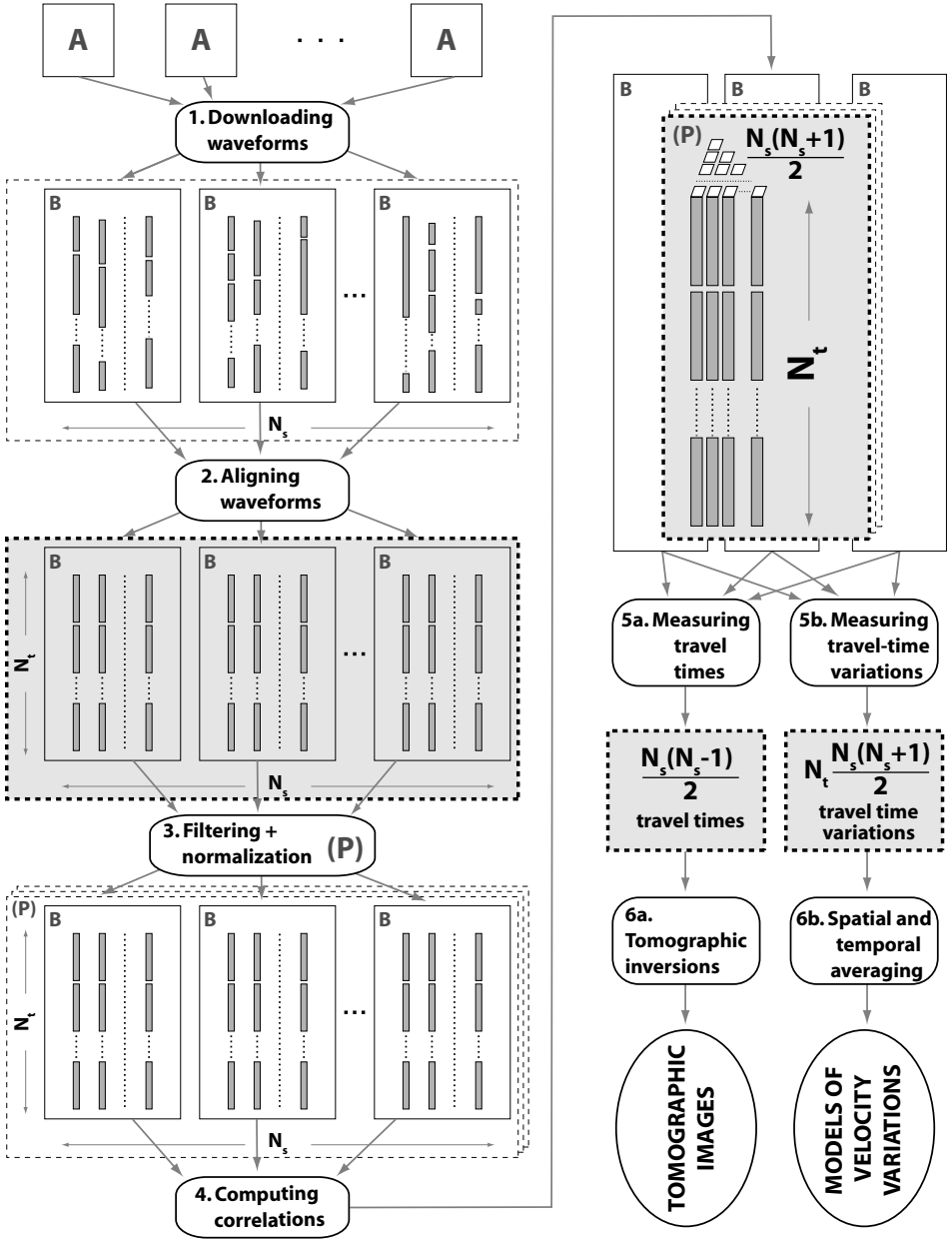


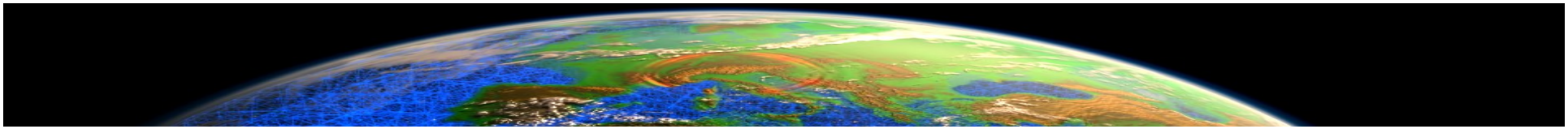
Brenguier *et al.* (2008)

Noise-based seismic tomography: Application to Alps



Stehly *et al.* (2009)





Data Intensive analysis

Intrinsic infrastructure mismatch

- Data volumes increase 100x in 10 years
- I/O bandwidth improves ~3x in 10 years

Distributed Data Mining:

- **Distributed Mining** of Data (moving algorithms and operate with data in situ)
- Mining of **Distributed Data** (hierarchy of data storage, data reuse management)

Data Architecture:

- Archiving and access (from preservation to scientific exploitation) -> **database model**
- Mining and data analysis (indexing, and associative memory techniques, rapid query and search algorithms…) -> **scientific databases & data-centric architectures**

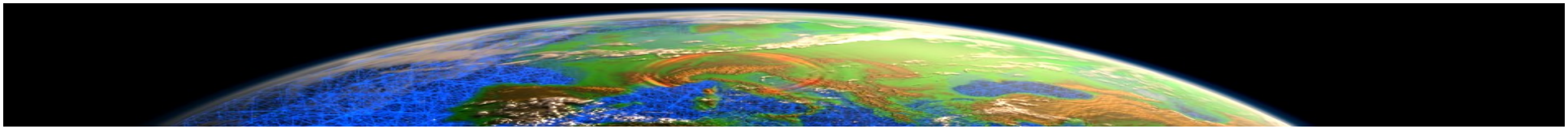
Amdahl's parallelism law: *when computation has a serial part S and a parallel component P, the maximum speedup is S/(S+P).*

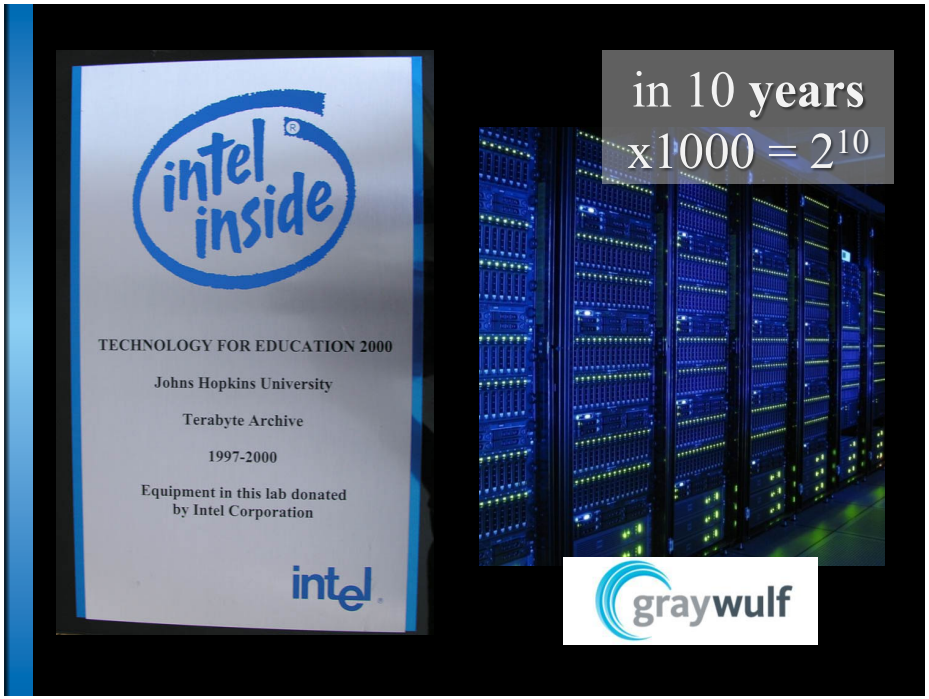
Amdahl's balanced system law: *A system needs one bit of IO per second per CPU cycle.*

Amdahl's memory law: *A system needs one byte of memory per CPU cycle*

Toward a data-centric environment:

- **Multiple tier model: hierarchical spread of memory and disk storage**
- **Hadoop/MapReduce versus Scientific Databases**
- **Shared and reusable data flows and workflows components**
- **Provenance and recognition**

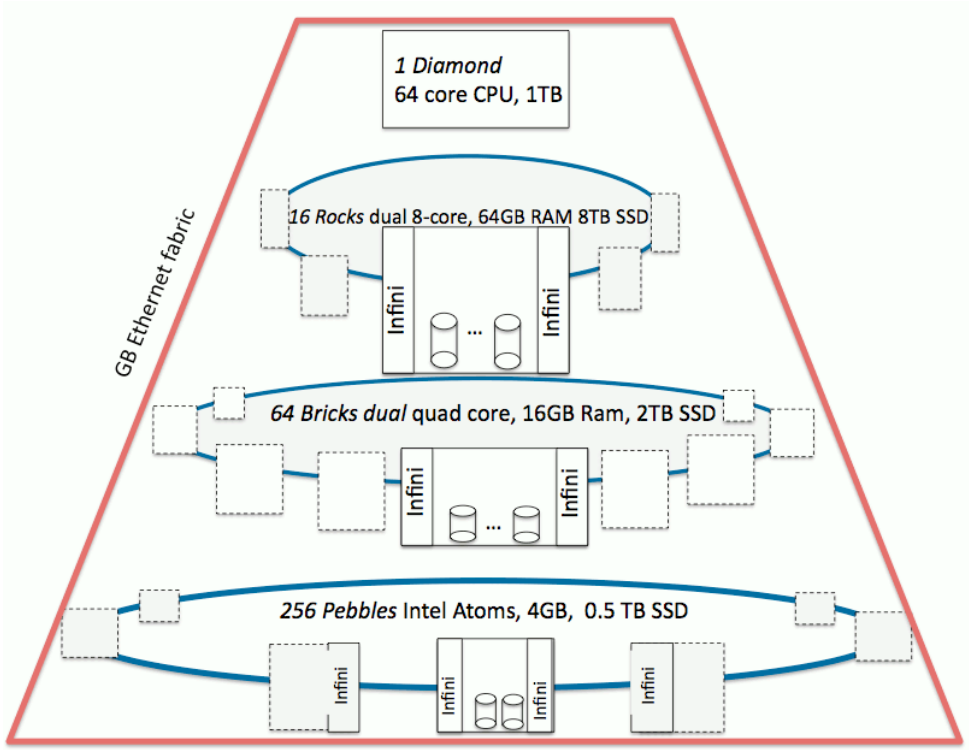


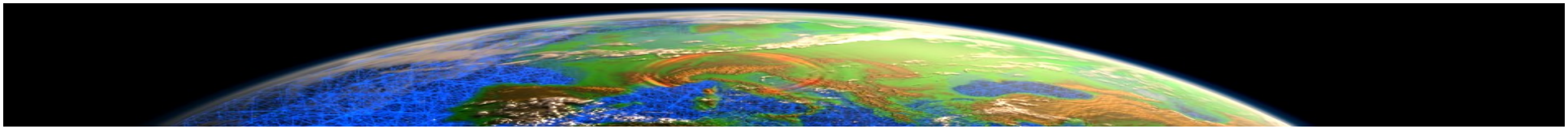


Sloan Digital Skye Survey

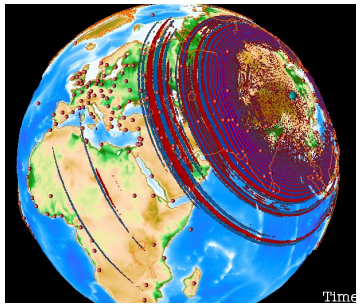
CWI platform

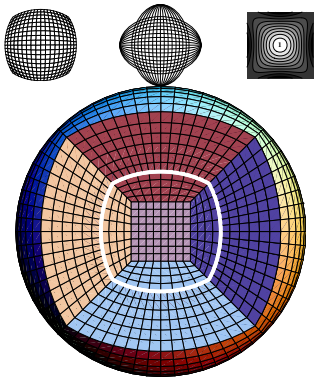
From Martin Kersten

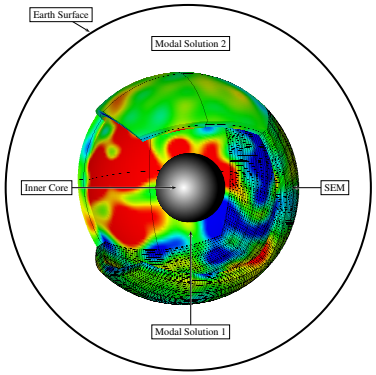




Data Intensive simulation and inversion







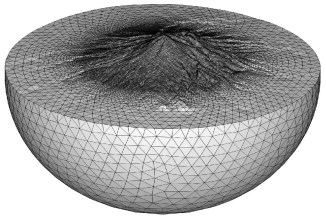
Global scale:

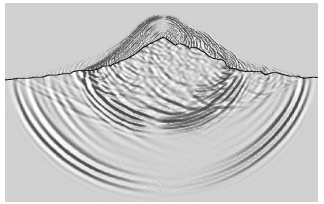
- **Synthetics prediction**
- **Full waveform inversion**
- NL inversion

Regional scale:

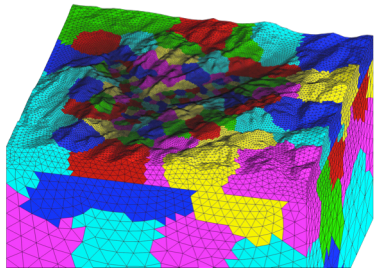
- **Complex wave propagation**
- Full waveform inversion
- **Extended earthquake sources**

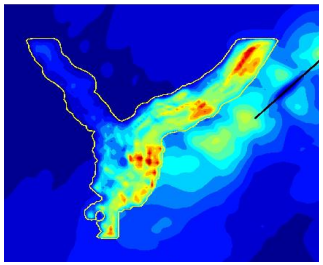
Aero-acoustic wave simulation in a volcano





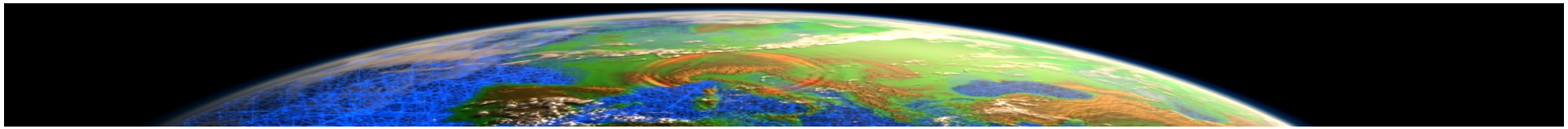
Strong motion simulation: Grenoble Valley





Strong motion prediction:

- **Synthetics prediction**
- **Earthquake source dynamics**
- **Stochastic simulation**



Data intensive computing challenges

Large scale 3D wave simulation:

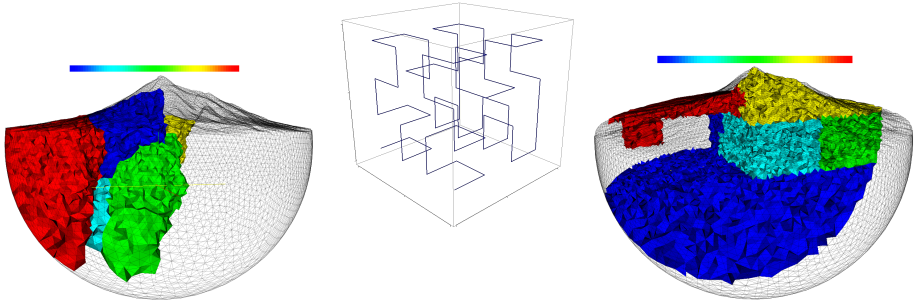
- synthetic seismograms
- Stochastic simulation

Full waveform inversion:

- adjoint inversion/optimisation methods
- Non linear probabilistic inversion

Orchestrated data analysis and data simulation

Hilbert SFC of level 2 and 64 sub-cubes



Domain decomposition by METIS (left) and SFC (right)

16 sub-domains, visible: 0-4

Scalability and Load balancing

Explicit locality: vertical and horizontal
Asynchronous high-level tasks concurrency
Scalable coordination and synchronization

Memory complexity

Memory node's hierarchy
Advanced data-structure
Scalable and efficient parallel I/O

Multicore architectures

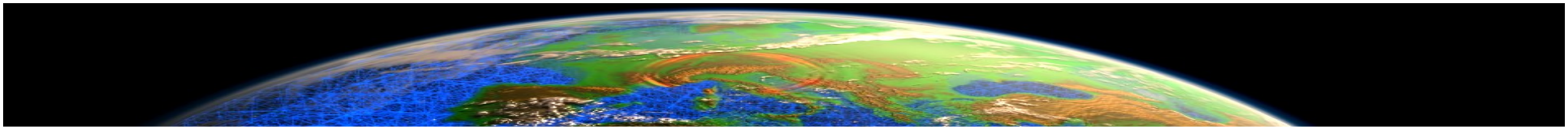
Mixed-hybrid parallel implementation
Self-scheduling at thread level
New instruction set node's architectures

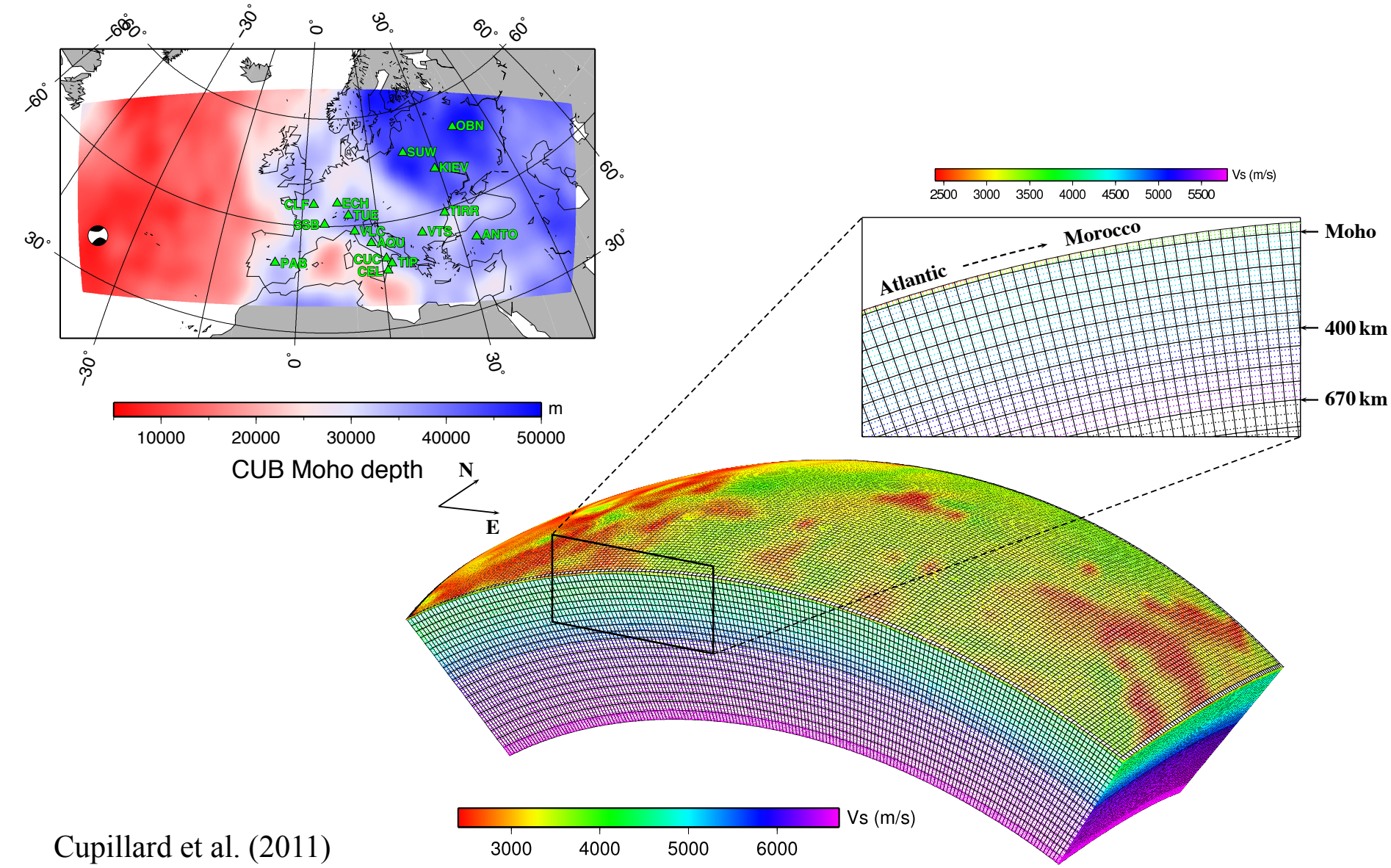
Orchestration workflows and visualization

Parallel unstructured mesh generation
Efficient adaptative domain decomposition
Seamless data movement across infrastructures

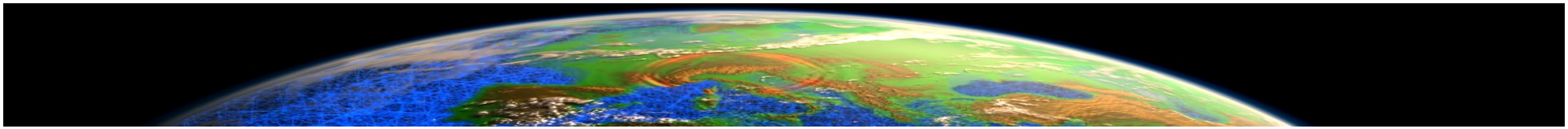
Community software building and sharing

Software engineering, refactoring…
Community code and community of practice ?

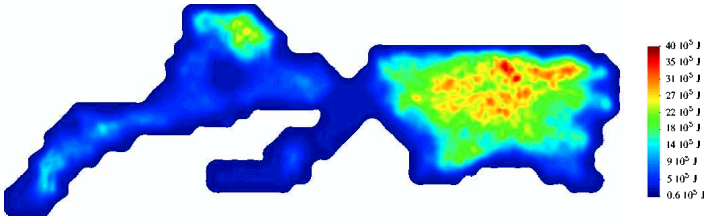




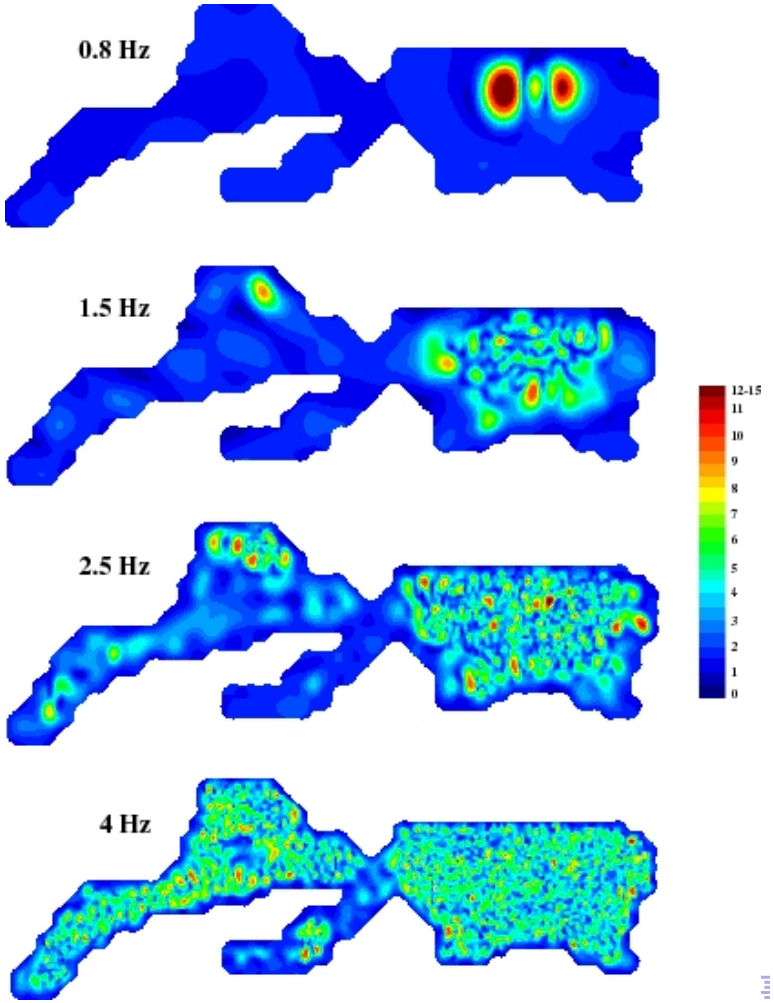
Cupillard et al. (2011)



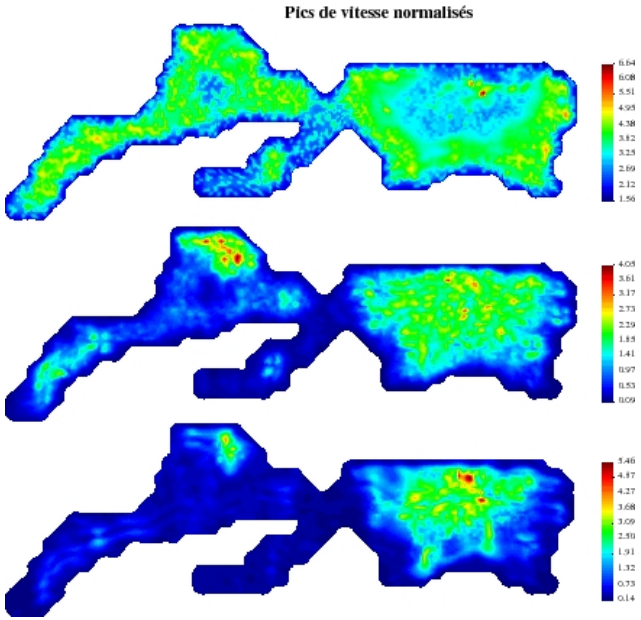
Kinetic energy



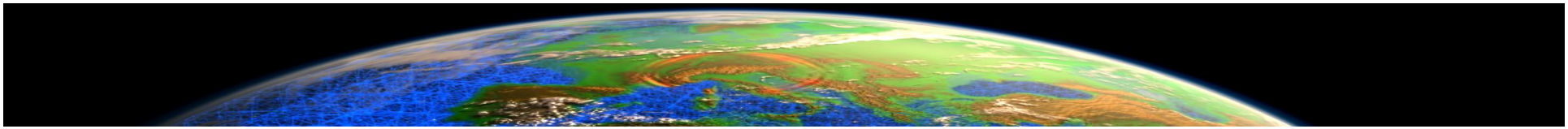
Transfer function



Normalized peak velocities



Delavaud et al (2006)



Towards an e-Science environment for seismology and EPOS

- **Provide a data intensive service-oriented e-Science environment to the EPOS community**
- **Lay the basis for transformative data-intensive research in the solid earth sciences**
- **Build trust and collaborative models for sharing of data , methods and tools**
- **Engage a new generation of researchers and experts in solid earth data intensive research**

European and International domain context

Integrated European distributed Data Archives (EIDA), part of the international FDSN

A number of coordinated European projects in seismology: NERA, SHARE, GEM, ERC WHISPER, ITN QUEST…

The European Plate Boundary Observation System (EPOS): the ESFRI-PP project

Active collaborations within the FDSN with US (IRIS-DMC), and Japan (JAMSTEX, NIED)

European e-Science context

Fast evolution of seismology services and applications

RapidSeis Portal for Accessing & Processing FDSN archives

European initiatives: EPOS, ENVRI and EUDAT

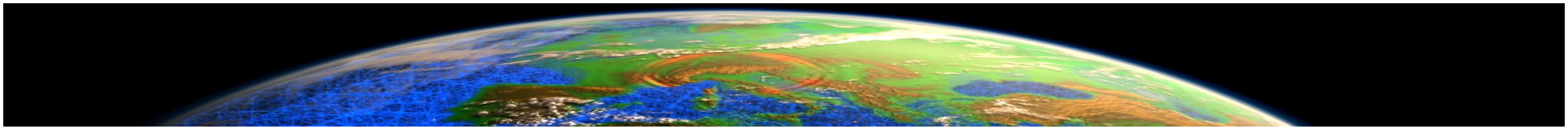
Converging e-Infrastructure ecosystem: EGI/NGIs, PRACE/NHPCs, GÉANT

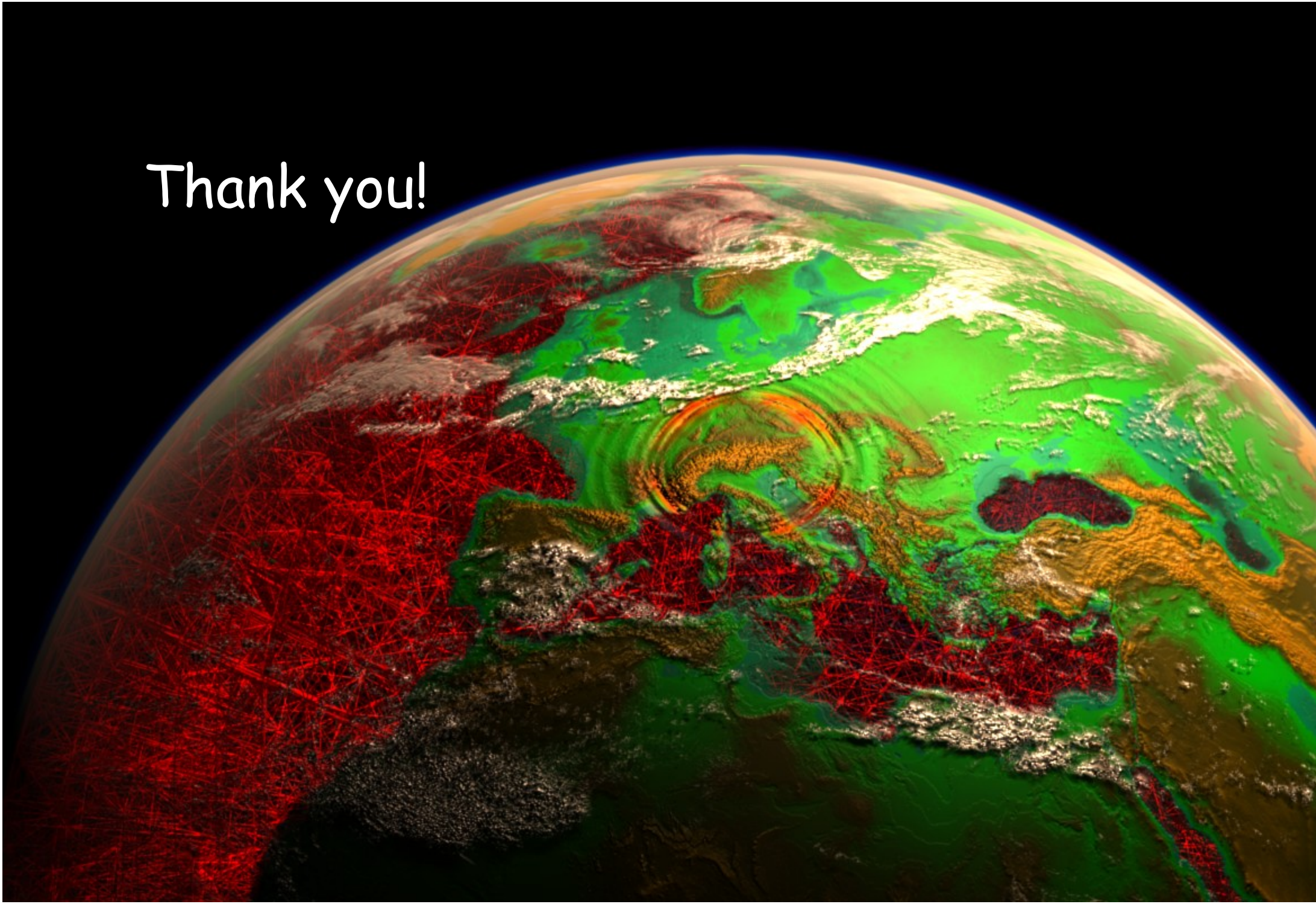
Emerging new data base management system and data centric architecture

Emergent data access and single-sign on protocols

A seismology architecture for data intensive applications: data analysis, mining and modelisation

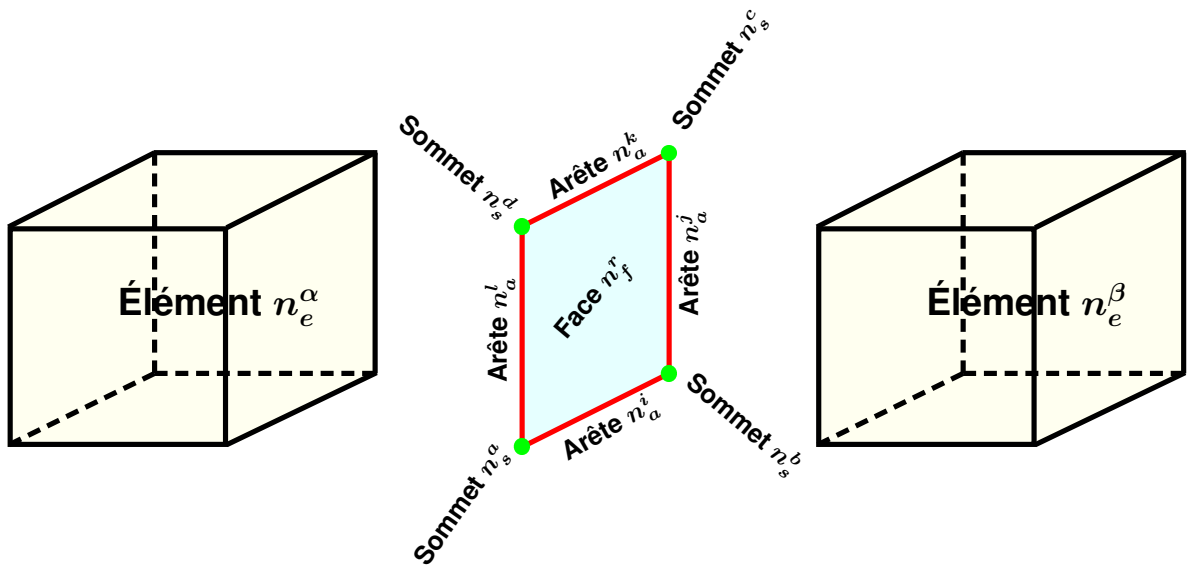
Sharing with other disciplines: Astronomy & Astrophysics, Particle Physics, Biology

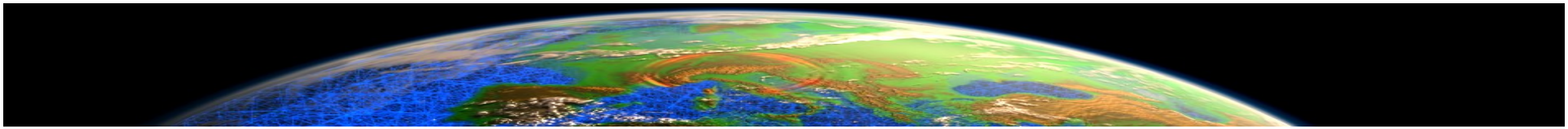



Thank you!

3D SEM code

- Fortran 90 syntax
- Parrallel (MPI) iimplementation – Domain decomposition **METIS**
- Unstructured mesh
- Object oriented





Data Intensive applications

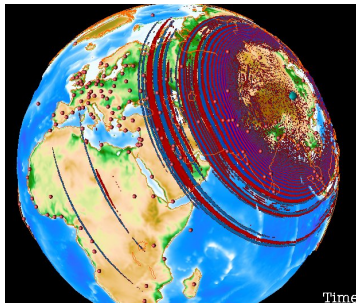
Earthquake and seismology community is facing a fundamental paradigm shift: from data driven to data intensive research:

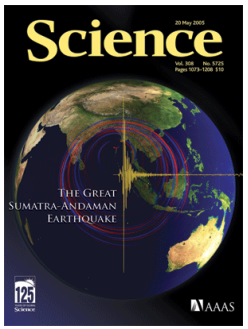
- Large volume data analysis: extracting information from space and time correlations in dense array observations,
- Data and computing intensive simulation/inversion: 3D wave form information using adjoint methods, stochastic strong motion simulation,
- Orchestrated workflows across service components.

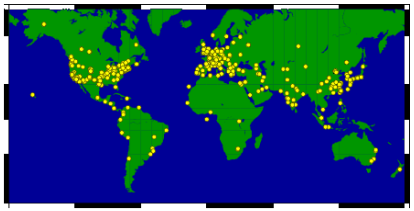
Seamless access to large volumes of multi-sets data across the Grid and HPC components

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)

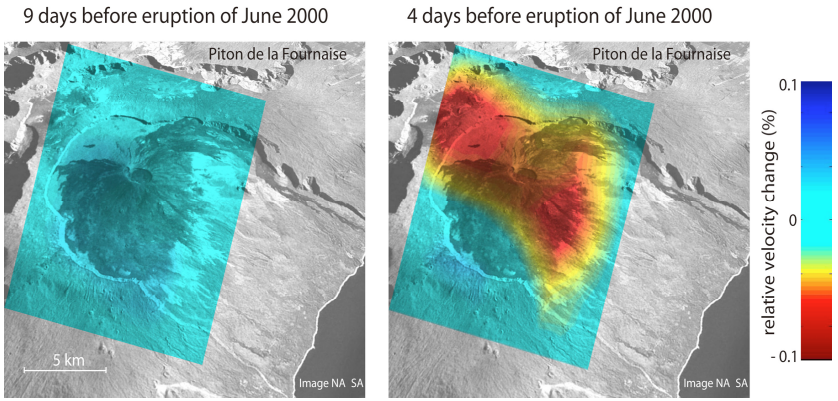




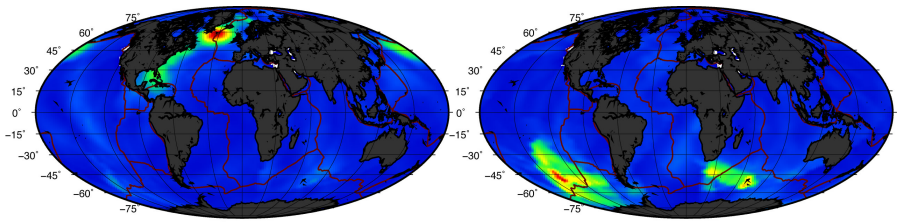


Research groups using SPECFEM3D

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