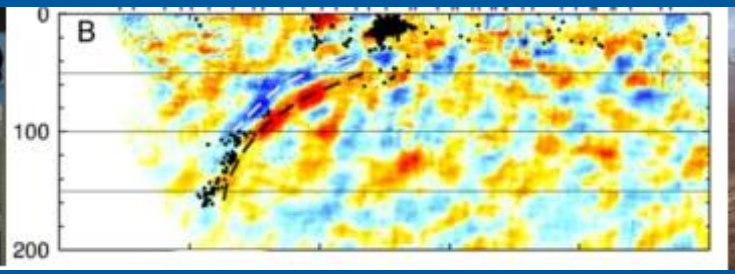
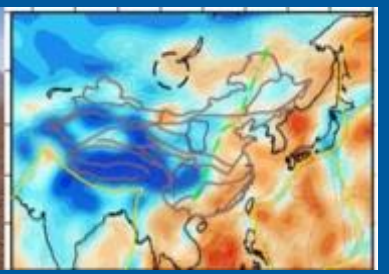
From seismic stations to integrated datacenters and computational facilities





Javier Quinteros, Angelo Strollo and the GEOFON team

The GEOFON mandate

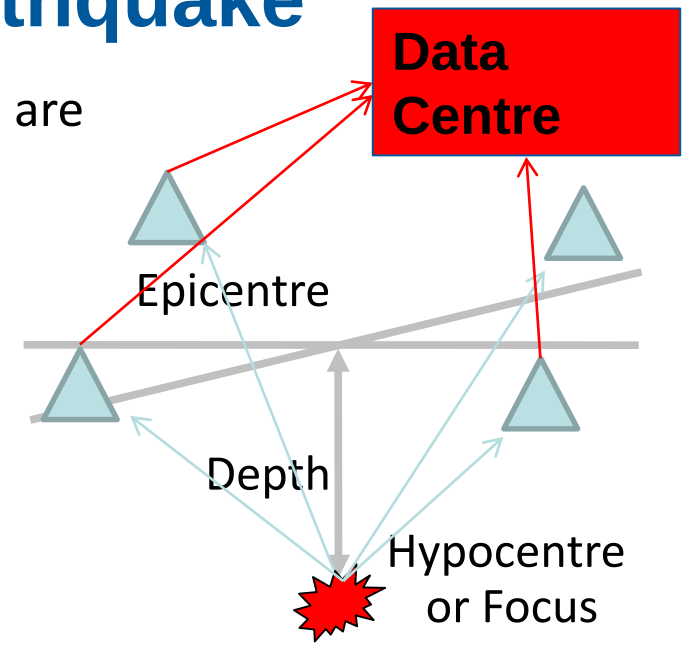
GEOFON aims at providing scientific infrastructure in the form of global seismic network, data archival and monitoring facilities to facilitate cooperation in seismological research leading to a better understanding of our complex system Earth.

It pursues these aims by operating and maintaining a global network of permanent broadband stations in cooperation with local partners, facilitating real time access to data from this network and those of many partner networks and plate boundary observatories, providing a permanent and secure archive for seismological data.

The GEOFON mandate

Using real-time data streams, GEOFON determines rapid automatic location estimates for all globally recorded earthquakes and most regional ones, and provides manually revised solutions for the largest earthquakes with minimal delay. This service provides basic rapid earthquake information to earthquake and tsunami warning centers worldwide, governmental agencies, disaster management teams, news media and scientists at the GFZ and elsewhere.

Describing an Earthquake

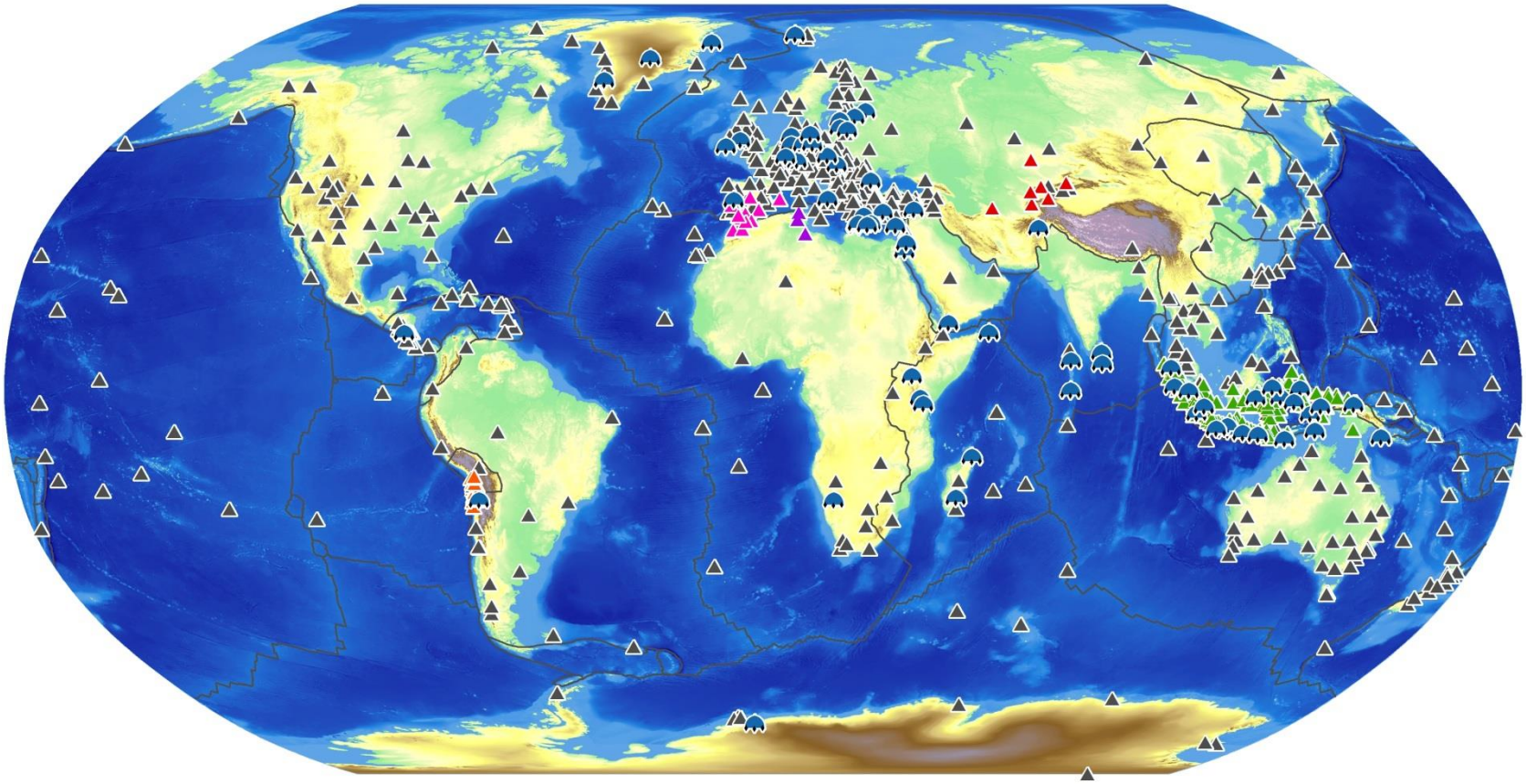


An earthquake's most fundamental characteristics are

- **Location**
  - ***Hypocentre*** or ***focus***: Geographical location of an earthquake i.e. its epicentre and depth.
  - ***Epicentre***: Point on the Earth's surface above the hypocentre.
- **Origin time**
  - Start time of the rupture.
- **Focal mechanisms**
  - ***Fault geometry:*** Geometric orientation of an earthquake's fault, dependent upon the local geology/tectonics.
  - ***First motion:*** Identifiable from the seismograms, used to identify the fault geometry
- **Size**
  - ***Magnitude*** or ***moment:*** Value describing the strength of an earthquake.

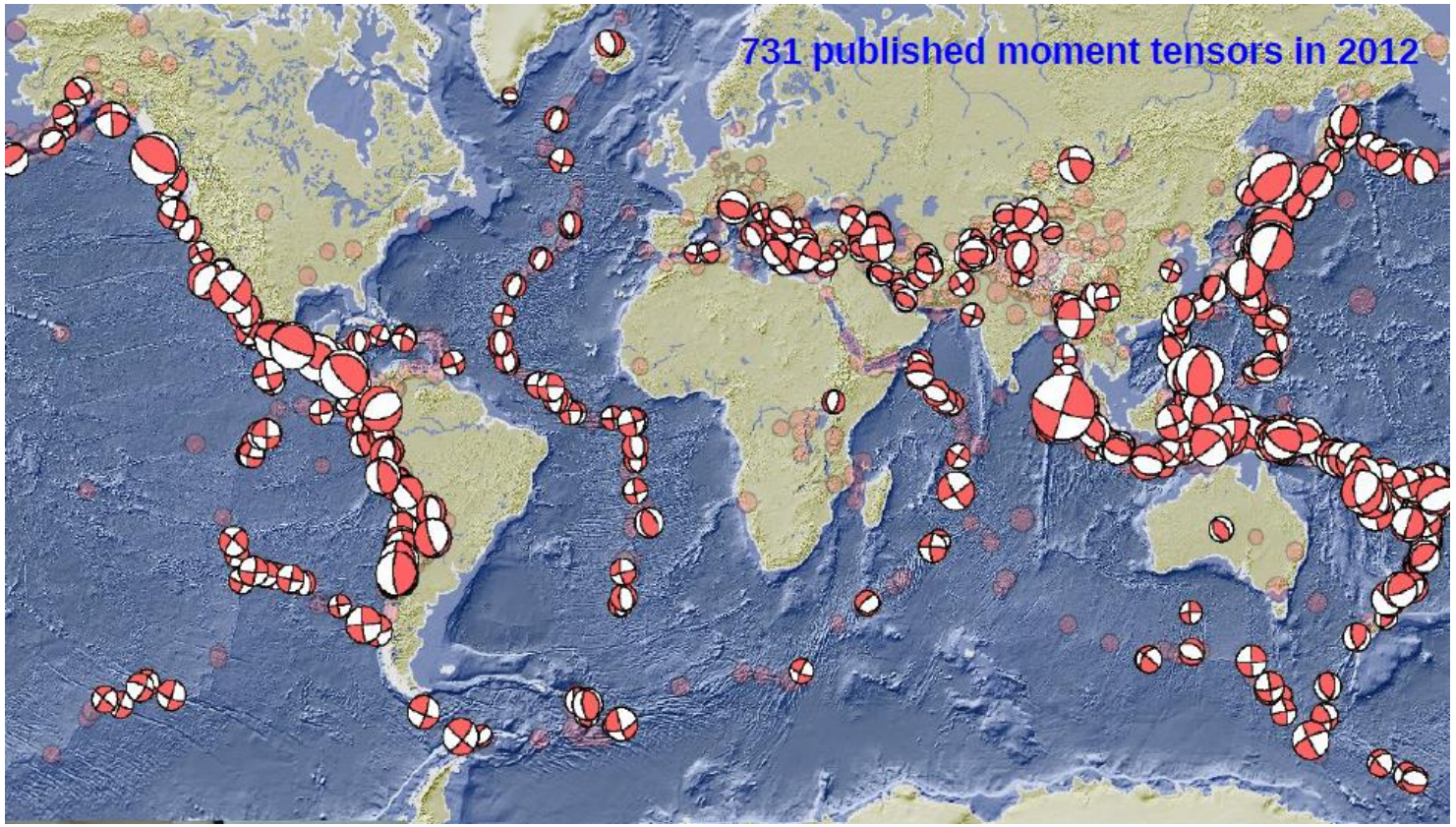
The GE global seismic network and GEVN

~80 GE stations, ~30 GFZ stations, ~800 stations for GEVN



Global earthquake monitoring

Since January 2011 also MT solutions for events larger than M 5 (depending on data quality)
6085 events and 731 MT solutions published in 2012



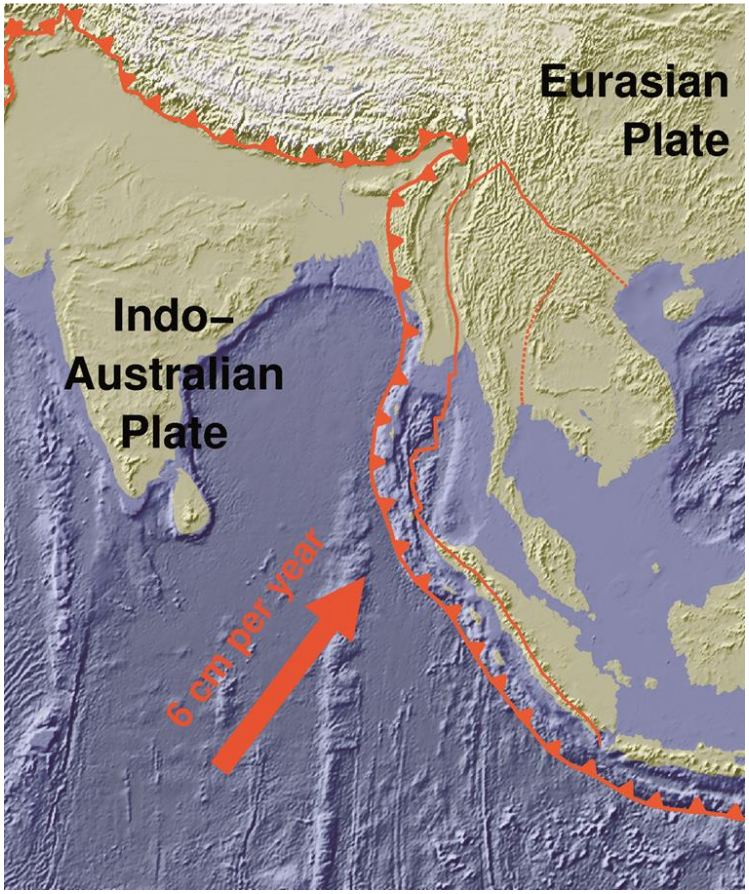
GEOFON in numbers

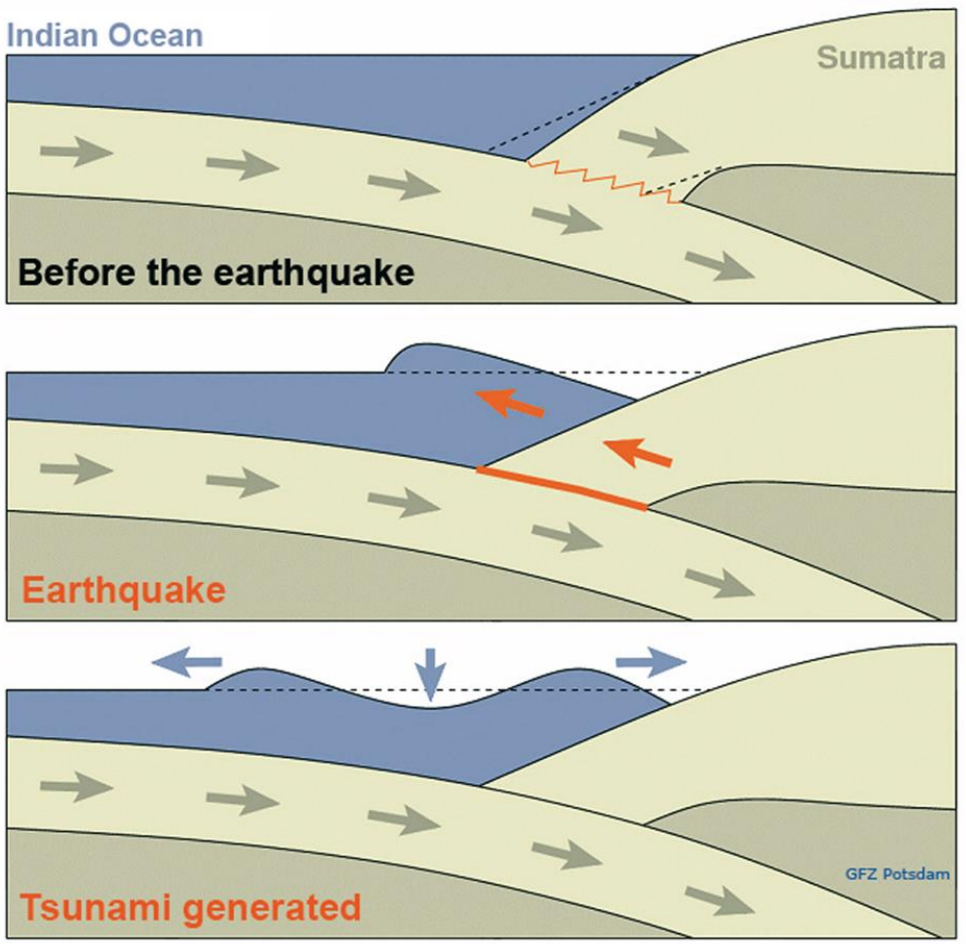
- **81** GE stations + **38** cooperated stations (affiliated)
- **~900** stations acquired and processed in real-time
- **~100** TB/year real-time data distributed via seedlink
- **~3000** stations **~70** TB in archive today
- **4.2** million customized requests in 2014
- **~6** TB/year of data shipped from archive
- **~6500** published events in a year
- **~1000** published moment tensors in a year
- **~28000**/day unique connections to the web pages (average 2014)
- **~330** SeisComP3 free non-commercial licenses

GEOFON data, products and software are open!

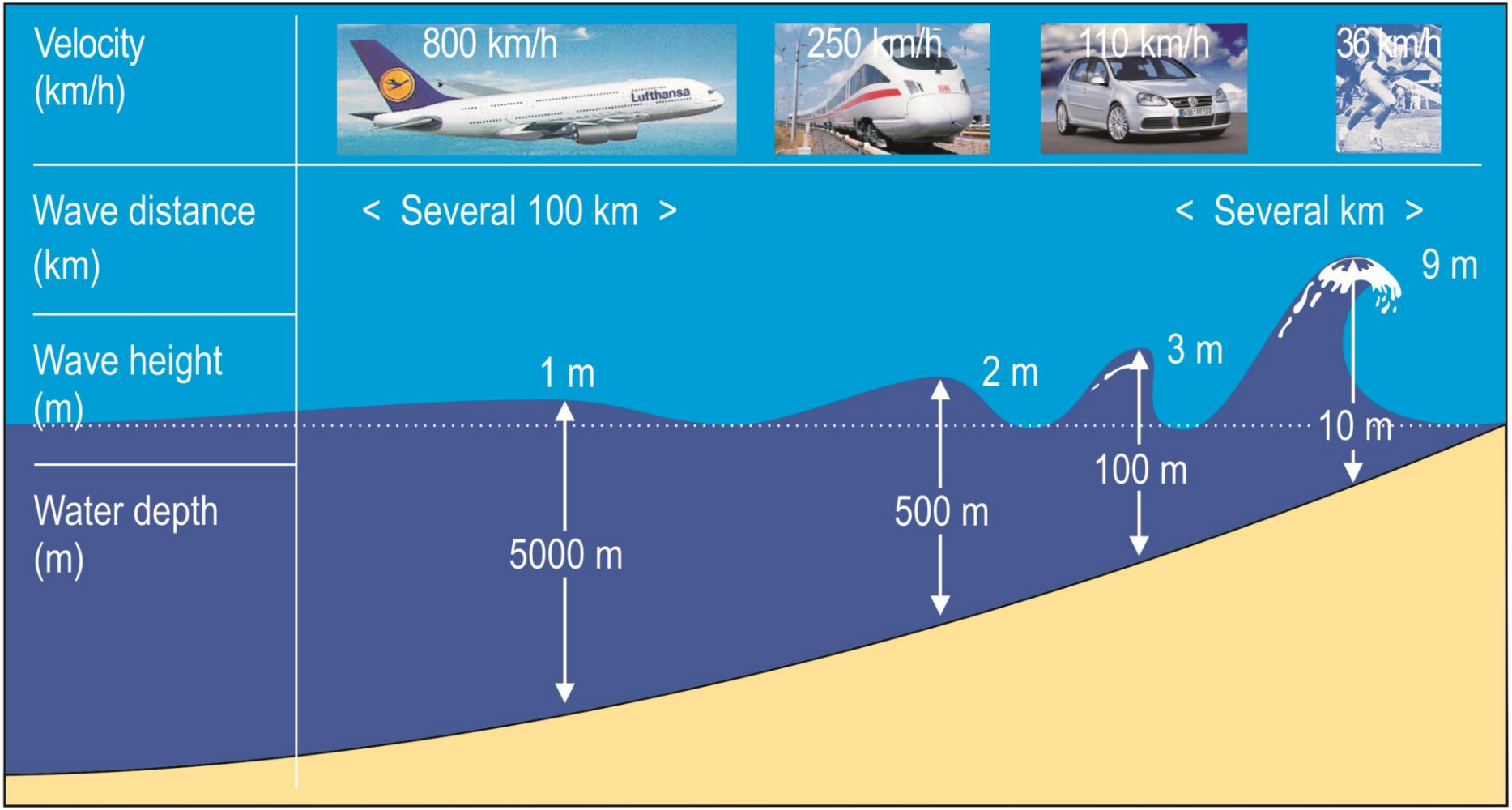
Early Warning Systems

Tsunami generation



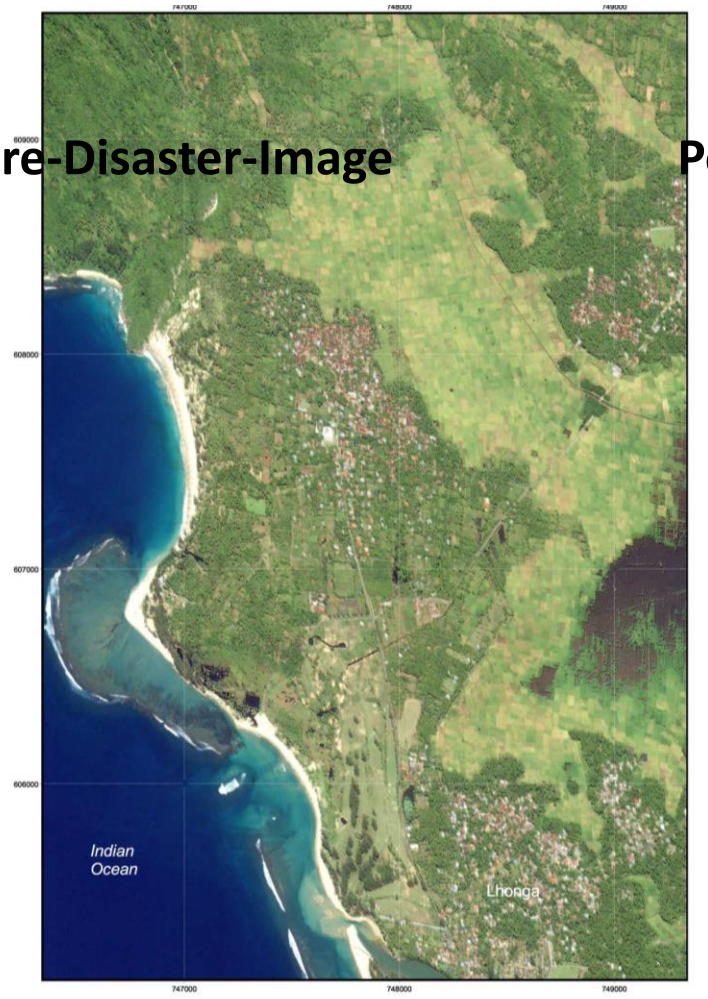


Tsunami propagation



The 26th December 2004 - Banda Aceh

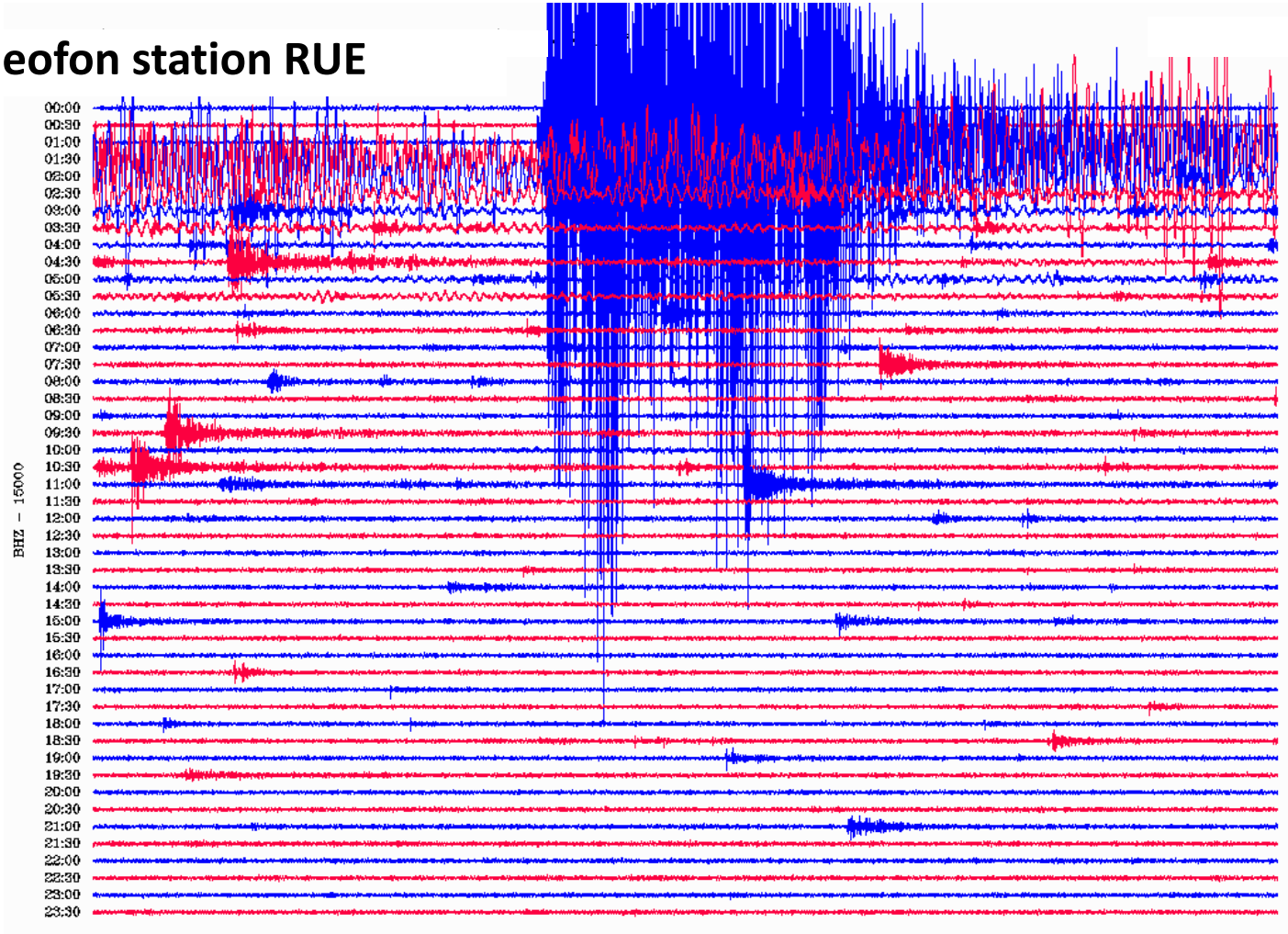
Pre-Disaster-Image



Post-Disaster-Image

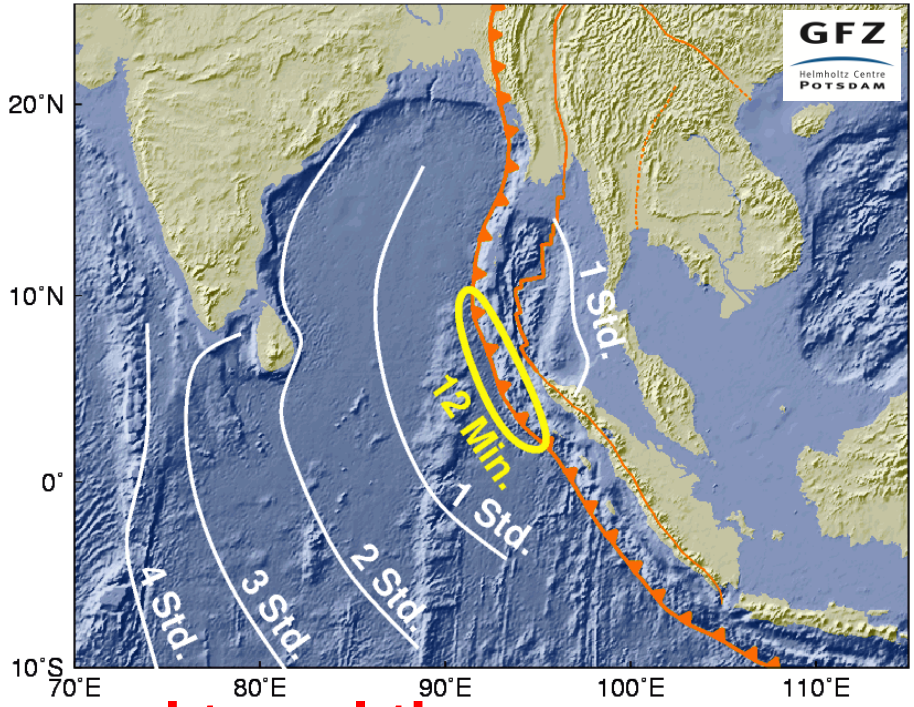
The 26th December 2004 - Germany

Geofon station RUE



The Challenge, Tsunami Warning for Indonesia

- **Earthquake sources very close to affected coasts**
- **Tsunami travel time 20 - 40 minutes**



- **Short tsunami travel times require tsunami warning within ≈ 5 minutes**

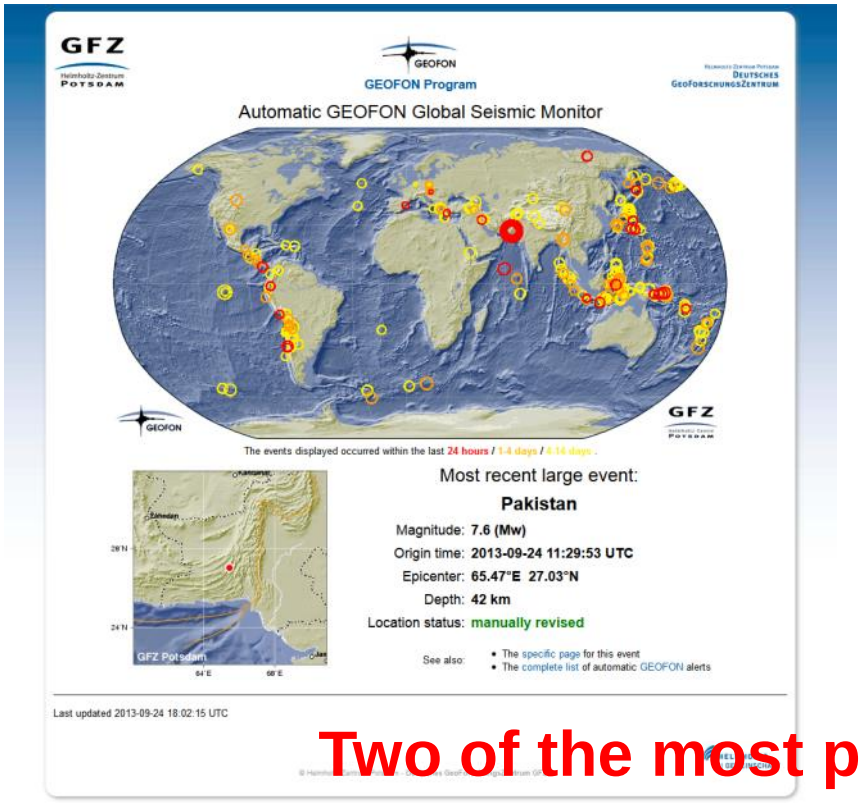
The 2014 Iquique Earthquake

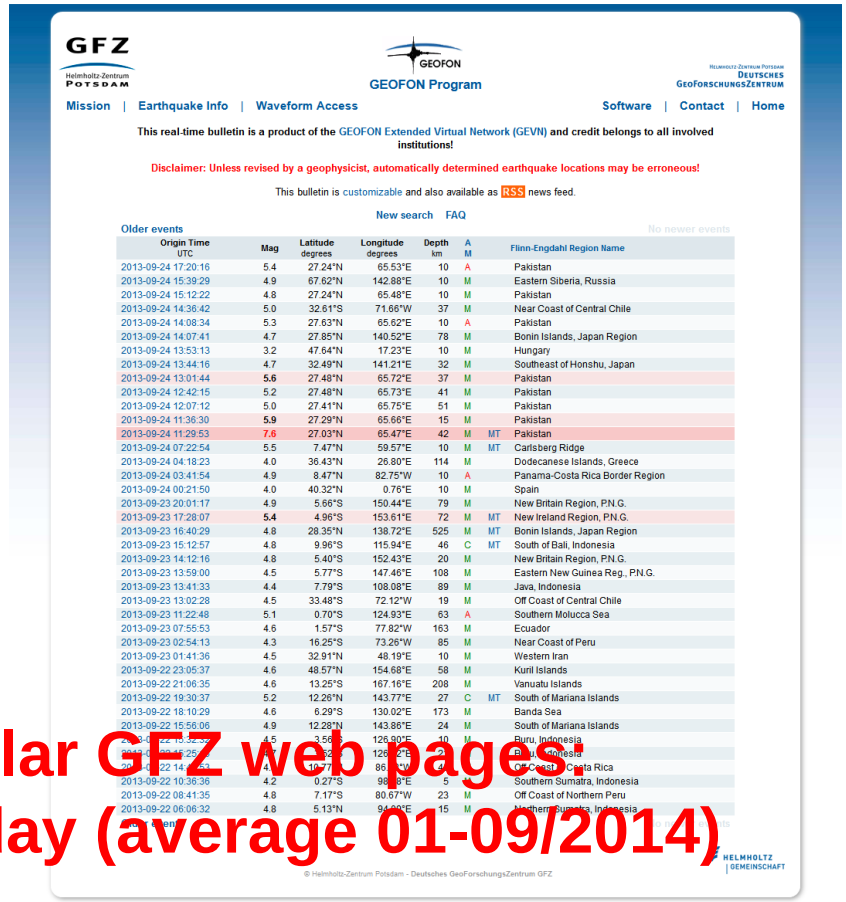
From rapid event information to scientific activities through HART deployment and data curation

Time

- 2014.04.01 23:46:49 UTC Event OT
- + 11 minutes - 1st automatic solution M 7.9
- + 55 minutes - 1st MT manual solution Mw 8.0
- + 1 day - MT manual review Mw 8.1
- + 1 day - Event special page (DOI)
- + 1 day - HART Coordination meeting
- + 10 days - Deploying HART instruments
- + 5 months - Nature paper and more
- + 12 months - DFG proposal submission
- + 16 months - HART data collected/archived
- + 18 months - DFG proposal accepted

Rapid Earthquake Information





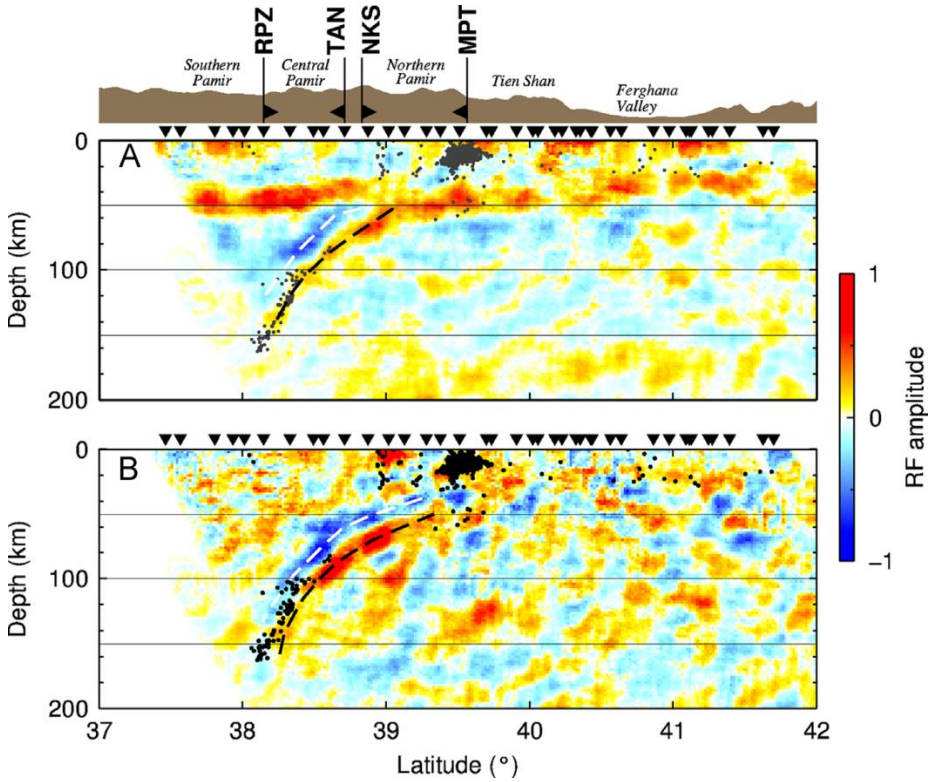
Two of the most popular GFZ web pages:
~28.000 unique visitors / day (average 01-09/2014)

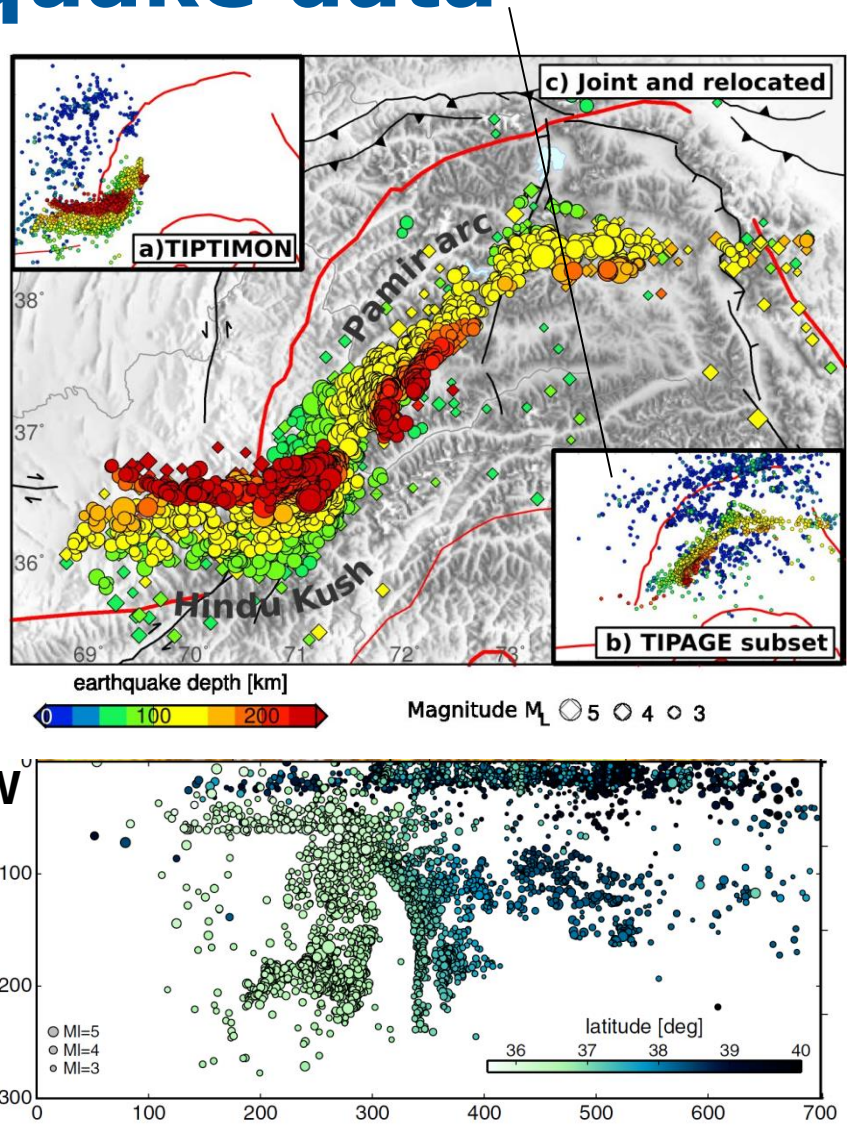
http://geofon.gfz-potsdam.de/eqinfo/seismon/globmon.php
http://geofon.gfz-potsdam.de/eqinfo/list.php

Scientific use of data

Imaging with earthquake data

- Moho, Lithosphere, Transition zone, Anisotropic structure
- Example: Pamir-Hindukush continental subduction (4 sections, 2 departments directly involved)

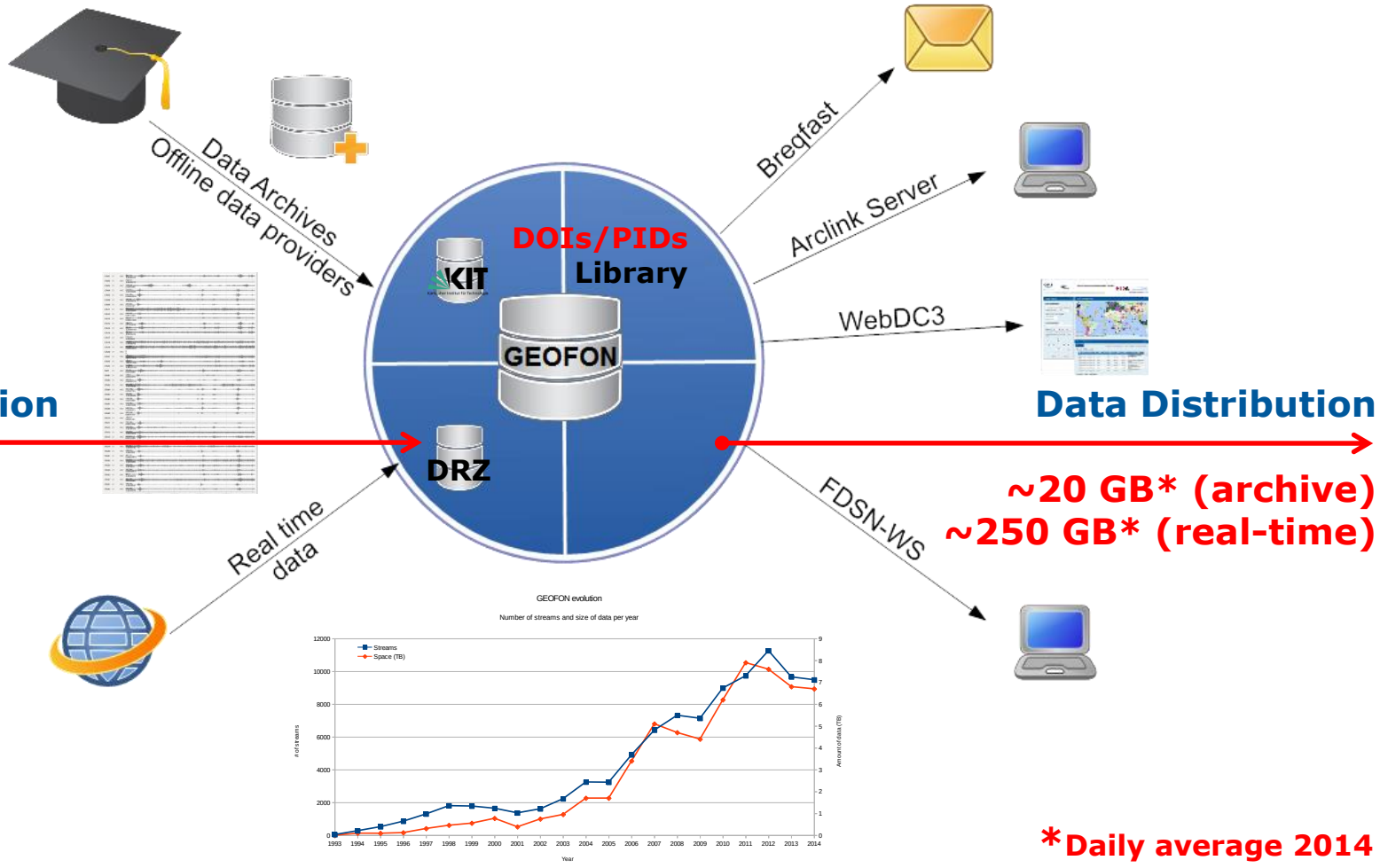




Schneider et al. (EPSL 2013), Sippl et al. (JGR, 2013), Kufner et al. (in prep)

GFZ scientific infrastructure

The GFZ seismological data archive: 70+ TB, 3000+ stations



*Daily average 2014

http://geofon.gfz-potsdam.de/waveform
http://eida.gfz-potsdam.de

EIDA/Orfeus
Construction of a European Federation

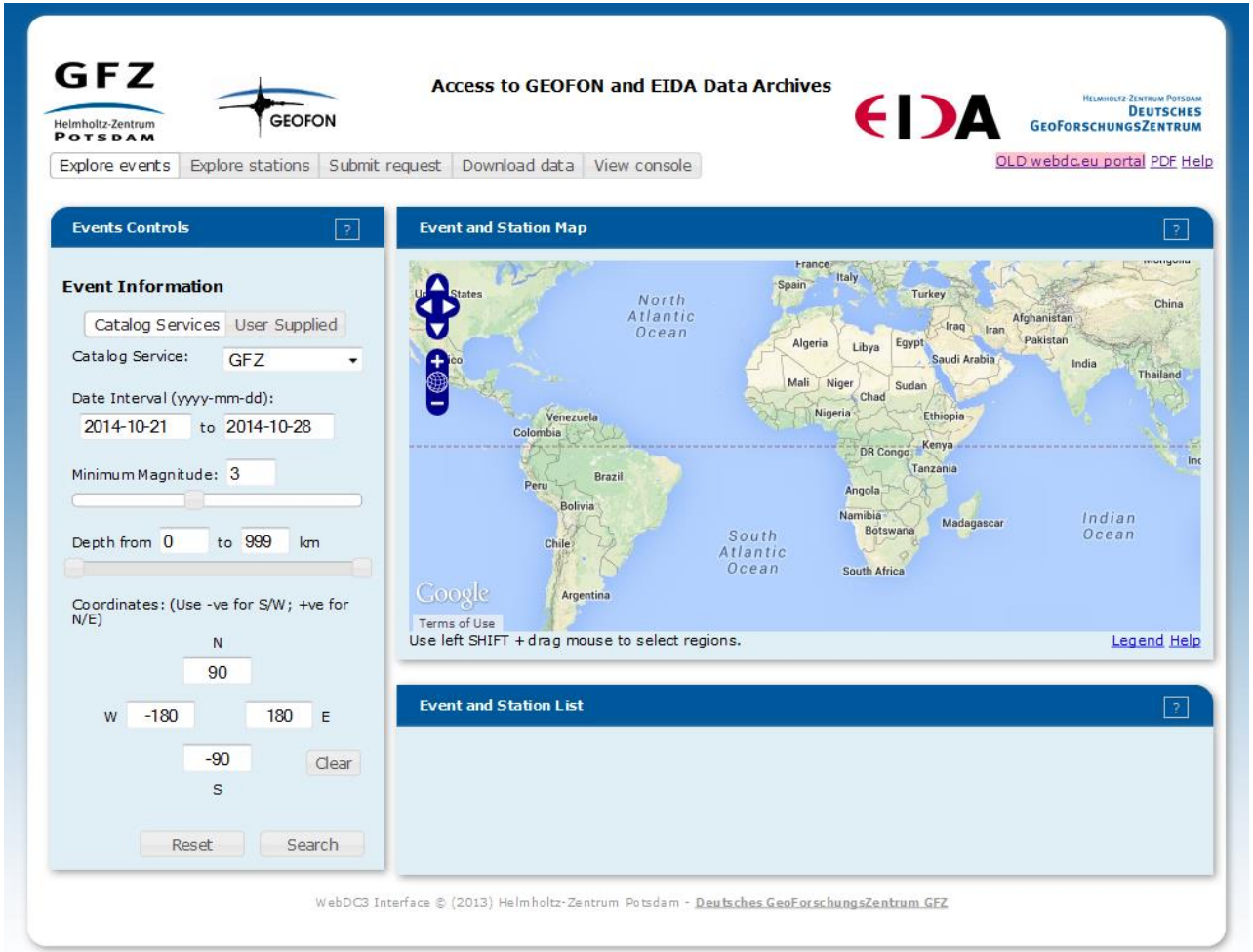
EIDA within Orfeus

EIDA within Orfeus

- Protocols to share data
  - Arclink protocol
- Synchronization of metadata
  - Daily
- Clients and tools development
  - SeisComP3
- Common policies for f.i. data curation
- Statistics
- Maintainance of Routing Tables

Seismic Waveform Data

How to get waveforms and metadata: WebDC3 portal



http://eida.gfz-potsdam.de

EIDA/Orfeus
New Projects: EUDAT, EPOS

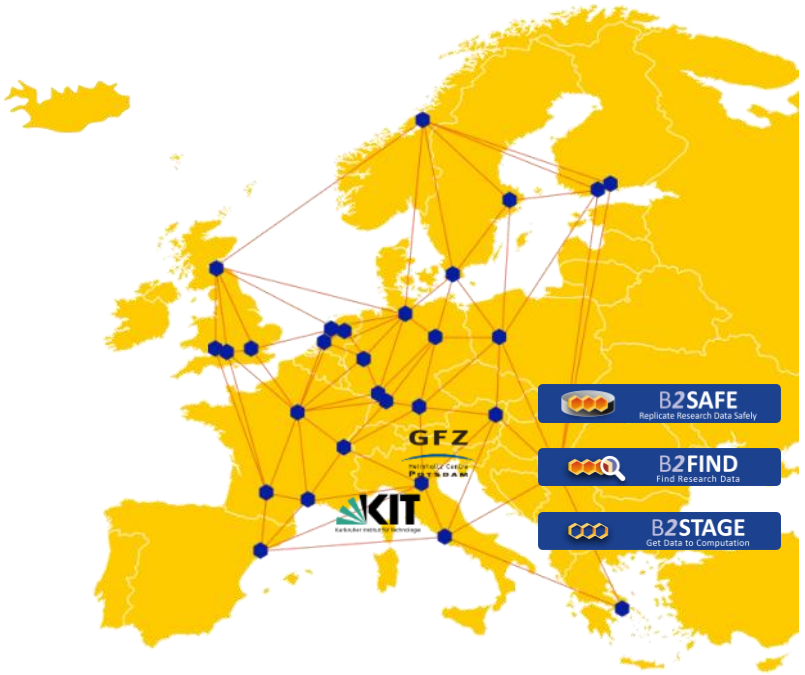
Seismic Waveform Data

EUDAT2020 The pan-European Data Infrastructure

EUDAT offers common data services, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network of 35 European organizations.

GEOFON involvement:

- **Safe replication**
- **Identity Management**
- **Data Discovery**
- **Dynamic Data**



EUDAT

http://eudat.eu/

Seismic Waveform Data

EPOS-IP European Plate Observing System

EPOS is integrating the diverse, but advanced European Research Infrastructures for solid Earth Science, and will build on new e-science opportunities to monitor and understand the dynamic and complex solid-Earth System.

GEOFON involvement:

- **EIDA-NG development**
- **Routing service**
- **Federated Identity management**

EUROPEANPLATEOBSERVINGSYSTEM

EIDA European Integrated Data Archive

http://www.epos-eu.org/

GFZ scientific infrastructure

Enhancing accessibility at European and Global scale

- GFZ seismological datasets are open data (except embargo for temporary experiments)

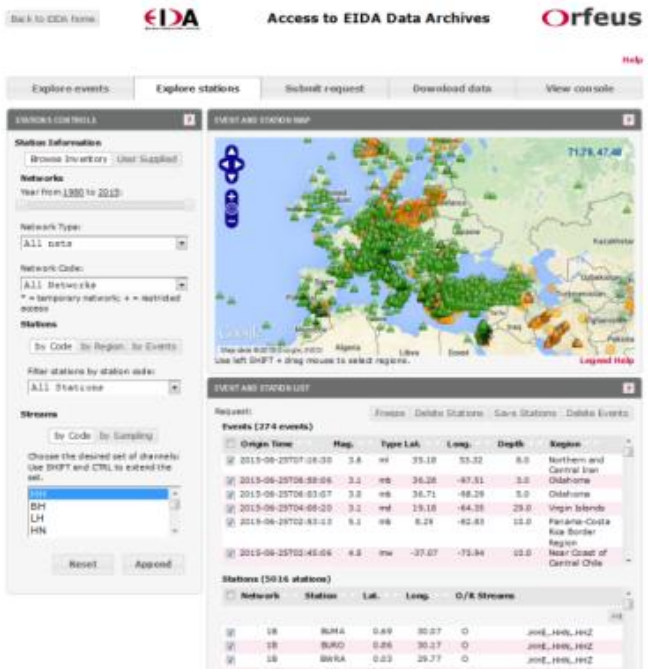
EIDA (European Integrated Data Archive)
=> **EUDAT** (The pan-European Data Infrastructure)
=> **EPOS** (European Plate Observing System)

Safe replication
Identity Management
Data Discovery
Data Staging
Dynamic Data
Integration and interoperability with other solid earth science infrastructure

EIDA Next Generation Software

More than just data and federated archive

- Coordination of data holdings and software/strategic developments
- Provides quality control of data/metadata
- Helps define seismological center best practice' for ORFEUS community



EIDA NG: Why?

Challenges:

- **Growth: archive volume, data variability, new nodes, more users, downloads**
- **More complex user requirements**

Requirements:

- **Scalability: beyond traditional seismology**
- **Interoperability / Compliance with standards**
- **Extensibility: services can be migrated to other communities**
- **Comprehensive handling of restricted datasets**
- **Extended search requests**
- **Combining QC / SoH information with data requests**
- **Data replication and identification: Serving user only the best datasets**

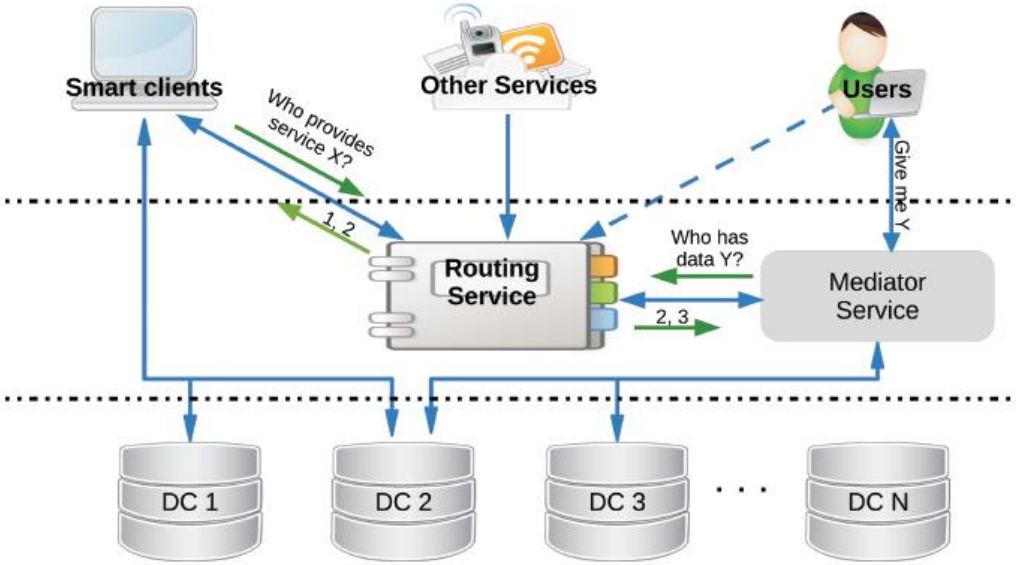
The EIDA Routing Service

What does a routing service?

- **Provides routing to data (streams)**
- **Routing to services**
- **Routing priorities**

Additional parameters information being discussed:

- **Geolocation of data and services**
- **Type of archive (master, validated/non validated copy, etc)**
- **Contact information**

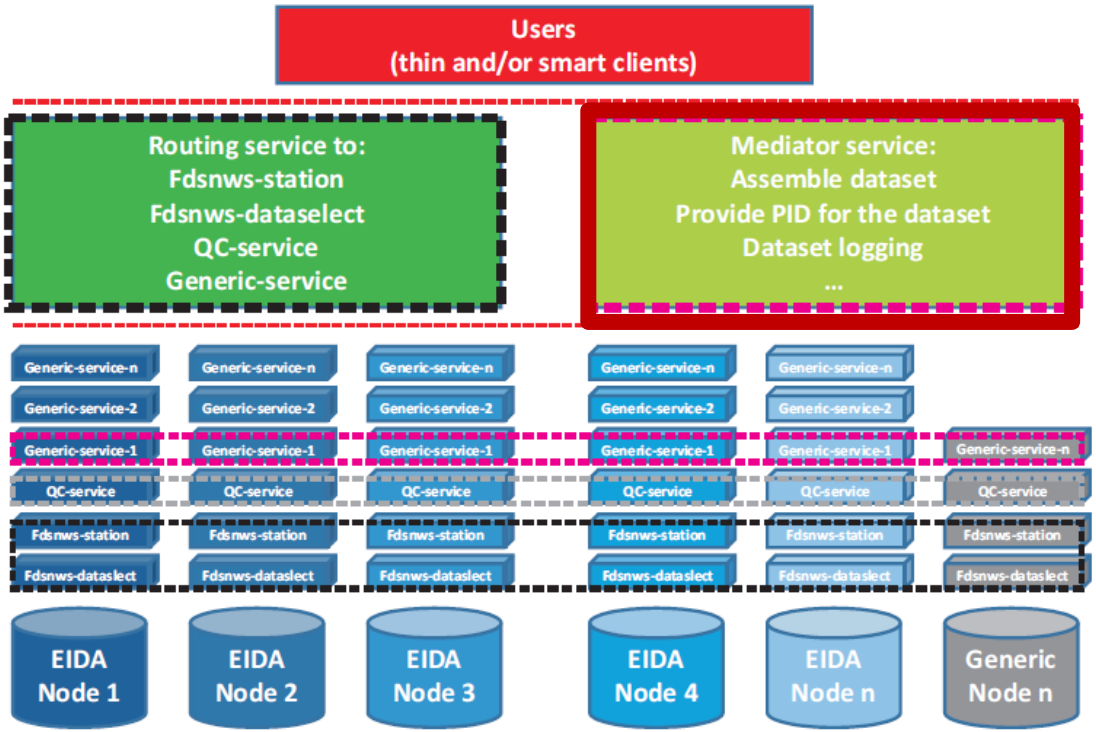


Can also be deployed as a standalone router to run on the client side to create virtual DCs

First stable release 10.2014

The EIDA Mediator (design phase)

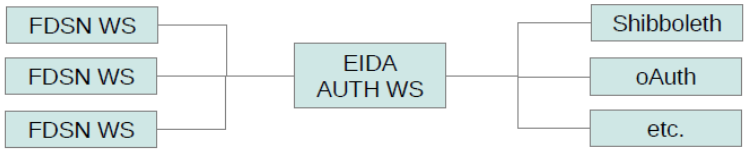
The Mediator will allow users to perform complex/filtered data requests based on a number of parameters available from a number of web services. Initially developed as a server/node side application, afterwards also as a client application.

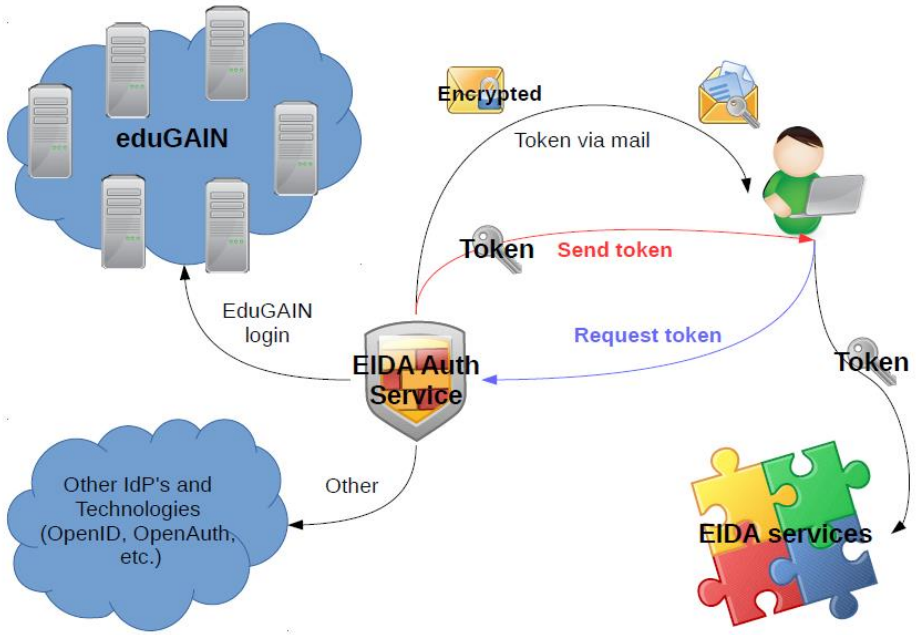


The EIDA AAI System (prototype)

Separate authentication from data services (leaving just authorization to data services).

- **Support multiple authentication mechanisms (Shibboleth oAuth, E-Mail, etc.).**
- **Secure authentication in scripts and browser.**
- **Pattern-based authorization (data access rules).**





„OLD“ Challenges from the seismological community

1: Standard Waveform Data Format
merging of different data sets
2: Standard (Real-Time) Data Exchange Protocol
Virtualization of Seismological Networks
3: Standard Data Archive Access Protocol
Virtualization of Seismological Data Archives
4: Synchronization of distributed data sets
5: Combining data and HPC services

Challenges from the seismological community

To be solved within EUDAT2020 and EPOS projects? (Hopefully!)

- Dynamic Data
  - ✓ Waveforms most probably will change only during the first days.
  - ✓ Seismology seems to be in a very good position compared to other sciences
- PID minting for waveforms
  - ✓ Time resolution? At the day level for every stream.
- Reproducibility of experiments
  - ✓ Freezing datasets in time
  - ✓ Millions of requests referencing thousands of records
  - ✓ Before and after curation of data/metadata
  - ✓ And real-time data?

Challenges from the seismological community

To be solved within EUDAT2020 and EPOS projects? (Hopefully!)

- Sharing datasets with the world
  - ✓ Same problems as the reproducibility
  - ✓ What about real-time data?
- Integration with other Earth sciences
  - ✓ Not enough standards in other branches of science
  - ✓ Different levels of evolution regarding infrastructure
  - ✓ Protocols
  - ✓ Data formats
  - ✓ Semantics! What does f.i. “deptht” means?

Thanks for your attention!