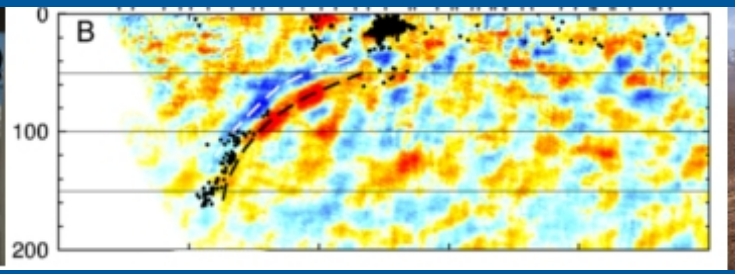
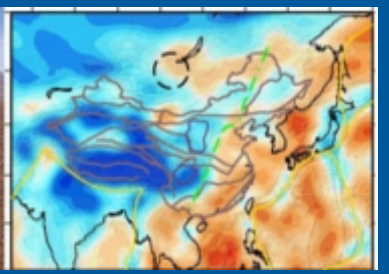
Seismology in 21st century From seismic stations to a professional Data Center



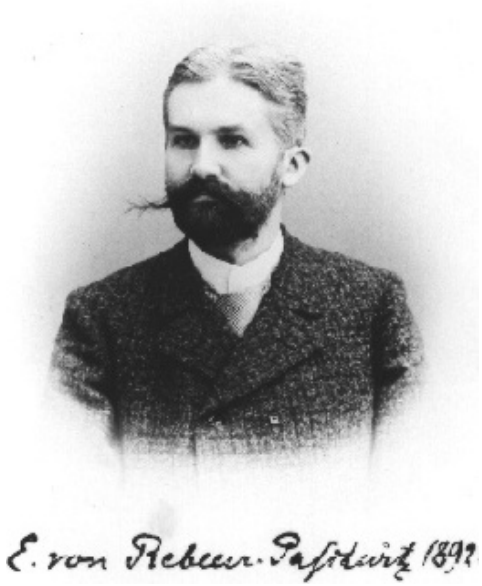


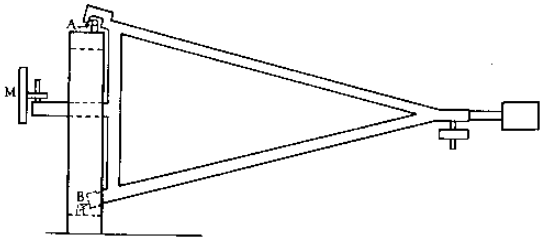
Javier Quinteros, Angelo Strollo and the GEOFON team

Outline

- The GEOFON Programm
- GEOFON Virtual Network
- Earthquake Monitoring
- Tsunami Early Warning System
- EIDA: A successful initiative of a virtual DC
- Latest developments: what can we expect in the future?

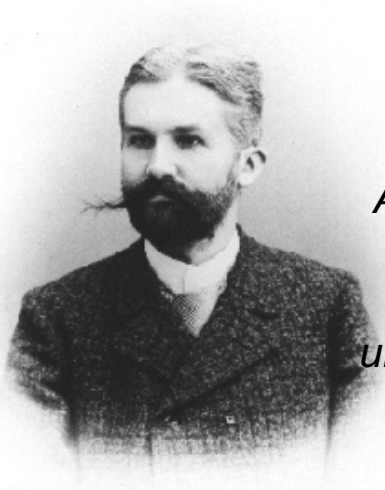
The first teleseismic recording





One of the first known recordings of a distant earthquake, obtained with von Rebeur's horizontal pendulum (*Nature*, **40**, 1889, p. 295). Two of these pendulums, located in Potsdam and Wilhelmshaven, recorded a large earthquake on **April 17, 1889**. The earthquake had been felt in Japan about an hour before it was recorded in Germany.

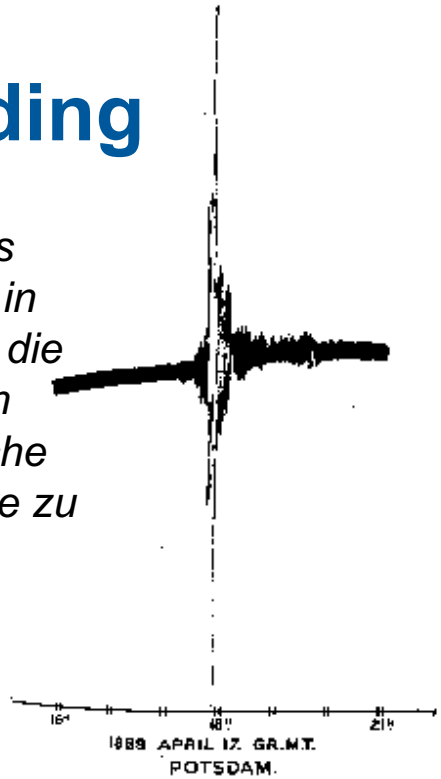
The first teleseismic recording



E. von Rebeur-Paschwitz 1892

"Wir wollen in erster Linie die Gründung eines internationalen Netzes von Erdbebenstationen in Anregung bringen, dessen Aufgabe es sein soll, die Ausbreitung der von großen Erdbebencentren ausgehenden Bewegungen auf der Erdoberfläche und durch den Erdkörper in systematischer Weise zu beobachten."

- Ernst von Rebeur-Paschwitz -



- **1992** GEOFON foundation. Permanent seismic Network + Data archive
- **1999** SeedLink real-time data exchange
- **2004** Start of Global Earthquake Monitoring
- **2004-06** ArcLink (WebDC) virtual data archive
- **2005-10** SeisComP3 and improved Global EQ Monitoring

- **2010** New GEOFON concept: Permanent and Virtual Networks, Global EQ Monitoring, Rapid Earthquake Information, European data archive, Networking and Capacity Building

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.

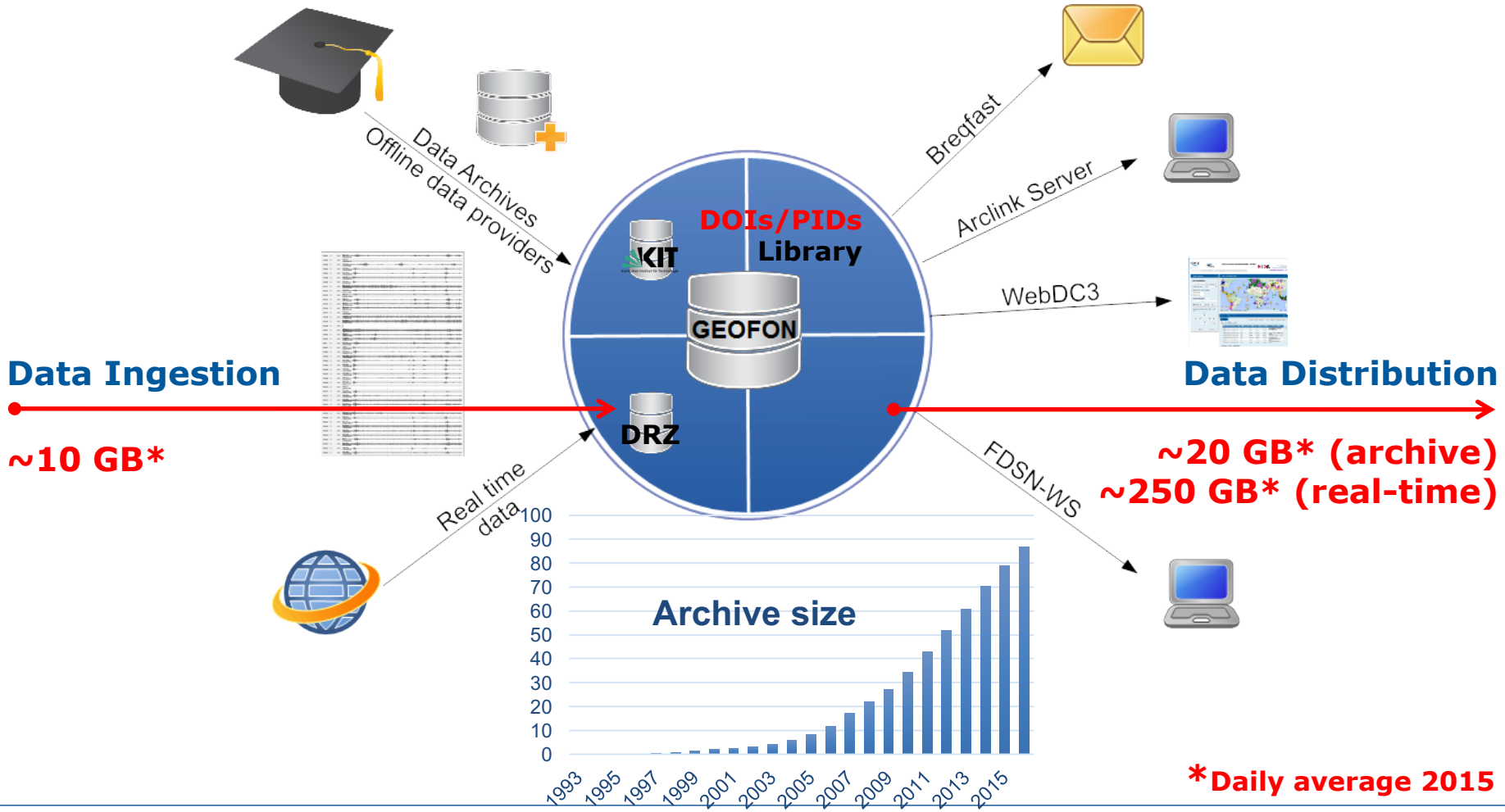
It pursues these aims by operating and maintaining a global network of permanent broadband stations in cooperation with local partners, facilitating real time access to data from this network and those of many partner networks and plate boundary observatories, providing a permanent and secure archive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

GFZ scientific infrastructure

The GFZ seismological data archive: +75 TB, +3000 stations



*Daily average 2015

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

GFZ Helmholtz Centre Potsdam

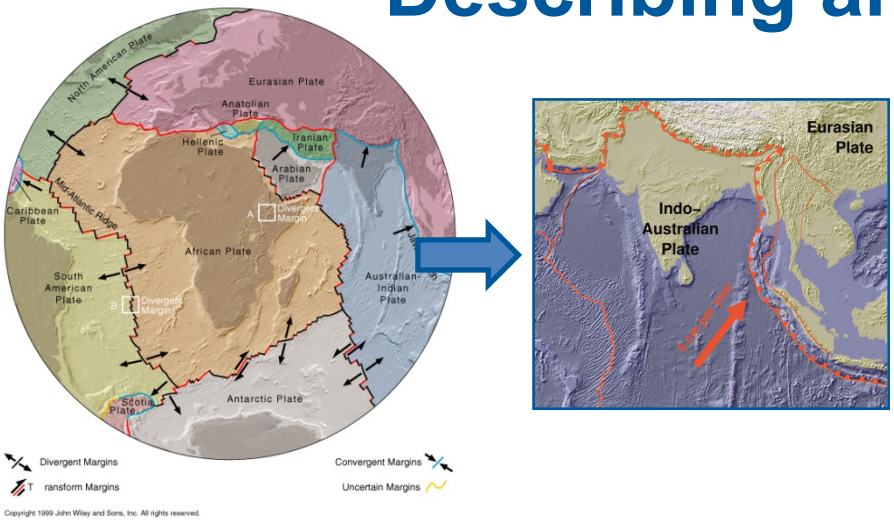
HELMHOLTZ ASSOCIATION

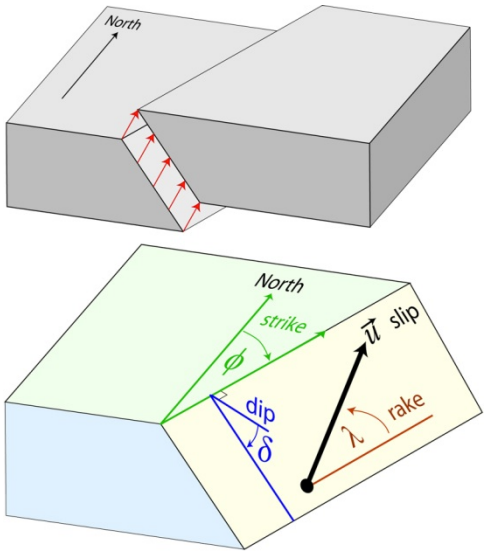
GEOFON in numbers

- **82** GE stations + **40** cooperated stations (affiliated)
- **~900** stations acquired and processed in real-time
- **~100** TB/year real-time data distributed via seedlink
- **~3000** stations **~75** TB in archive today
- **~4.5** Million customized (successful) requests/year
- +**200.000** notifications/year via e-mail or SMS
- **~6** TB/year of data shipped from archive
- **~6500** published events in a year
- **~1500** published moment tensors in a year
- **~28000** unique visitors/day to web pages (yearly avg.)
- **~370** SeisComP3 free non-commercial licenses

GEOFON data, products and software are open!

Describing an Earthquake

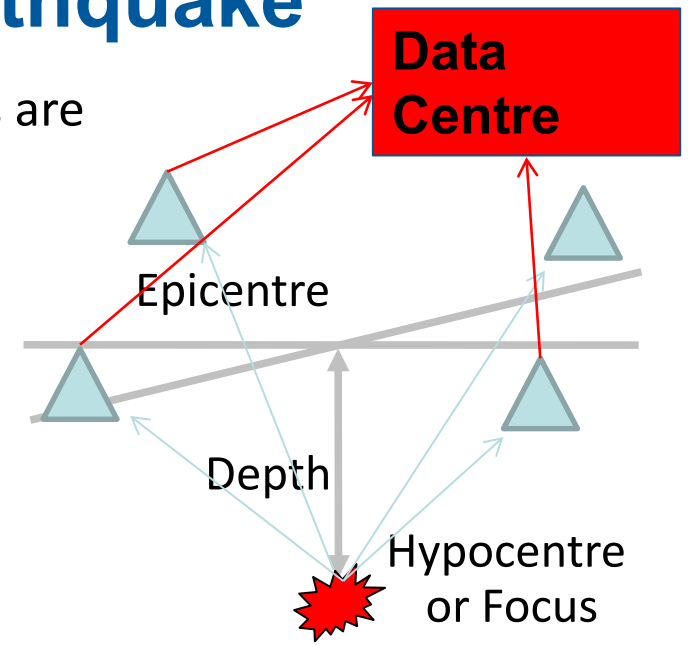




Tectonic earthquakes are caused by slippage along fractures (faults).

As slip along the faults occurs, energy is released which travels in the Earth interior in the form of seismic waves…

Describing an Earthquake

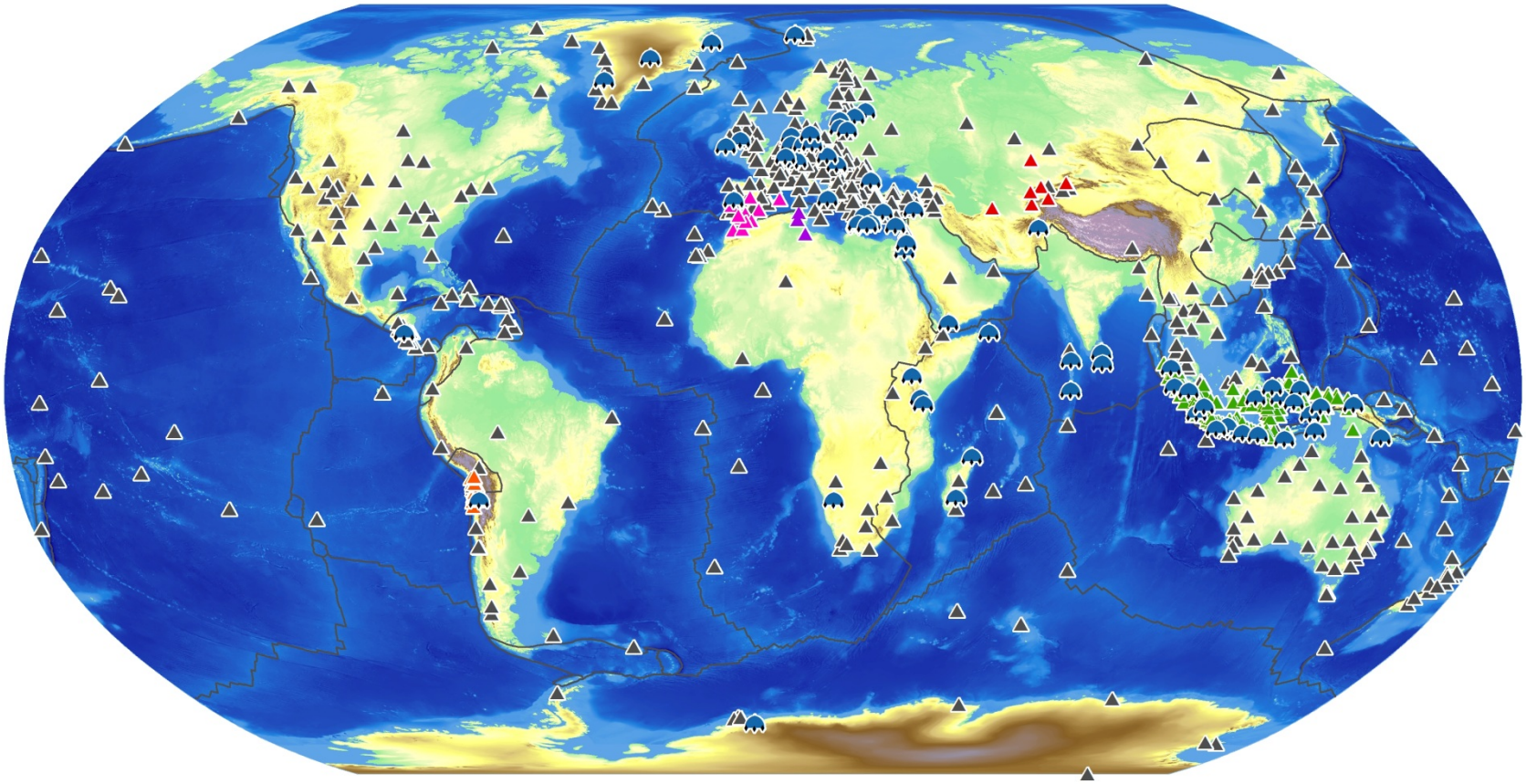


An earthquake's most fundamental characteristics are

- **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.

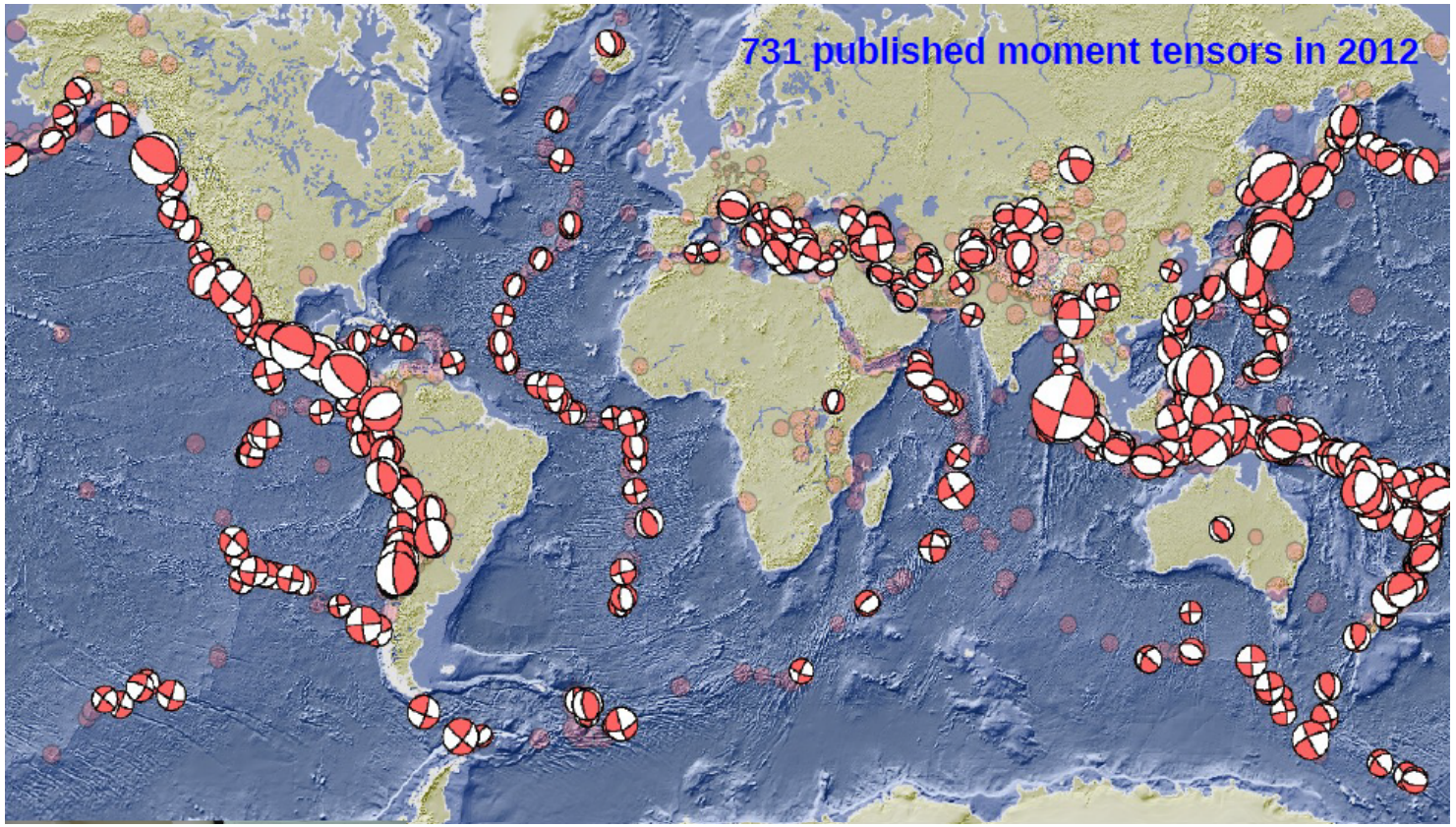
The GE global seismic network and GEVN

~82 GE stations, ~40 GFZ stations, ~900 stations for GEVN



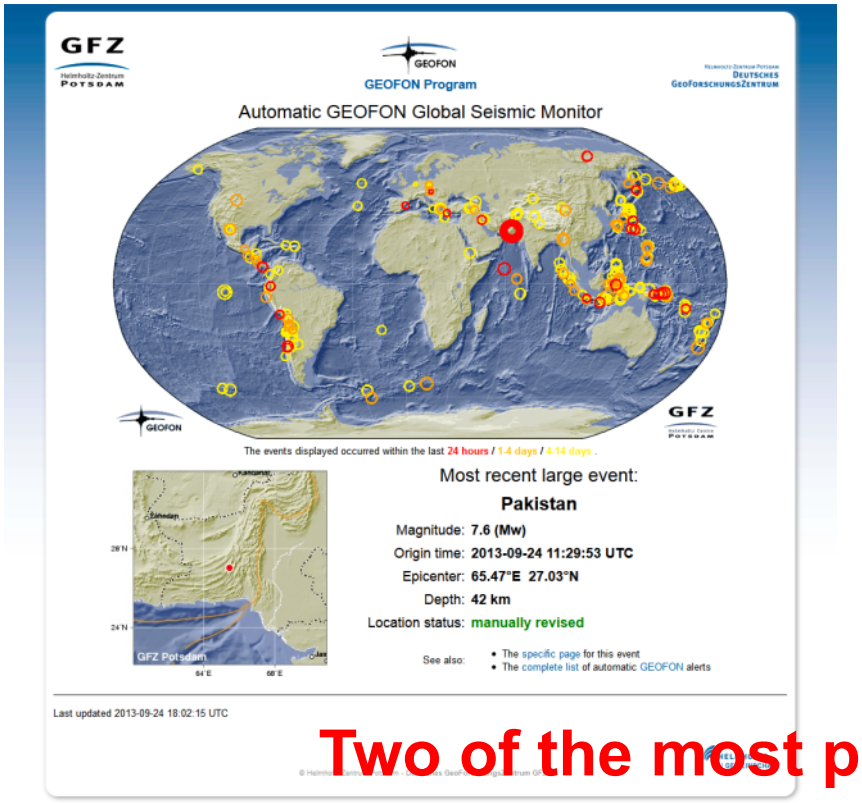
Global earthquake monitoring

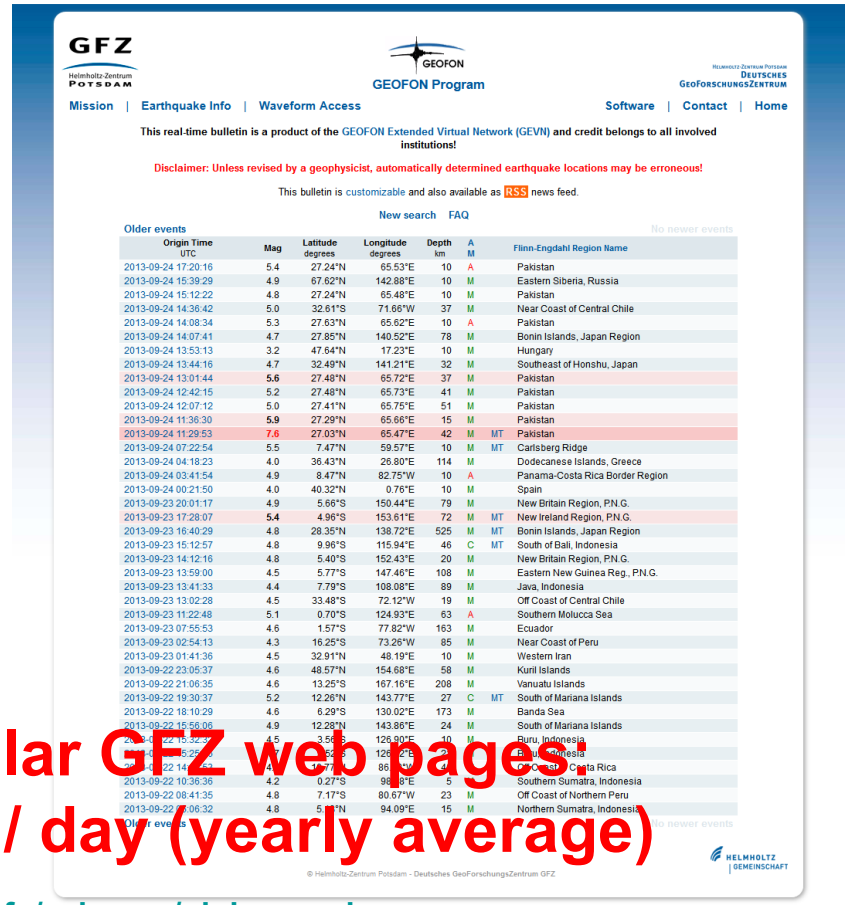
Since January 2011 also MT solutions for events larger than M 5 (depending on data quality)
~6500 events and ~1500 MT solutions published per year



Earthquake Monitoring

Rapid Earthquake Information





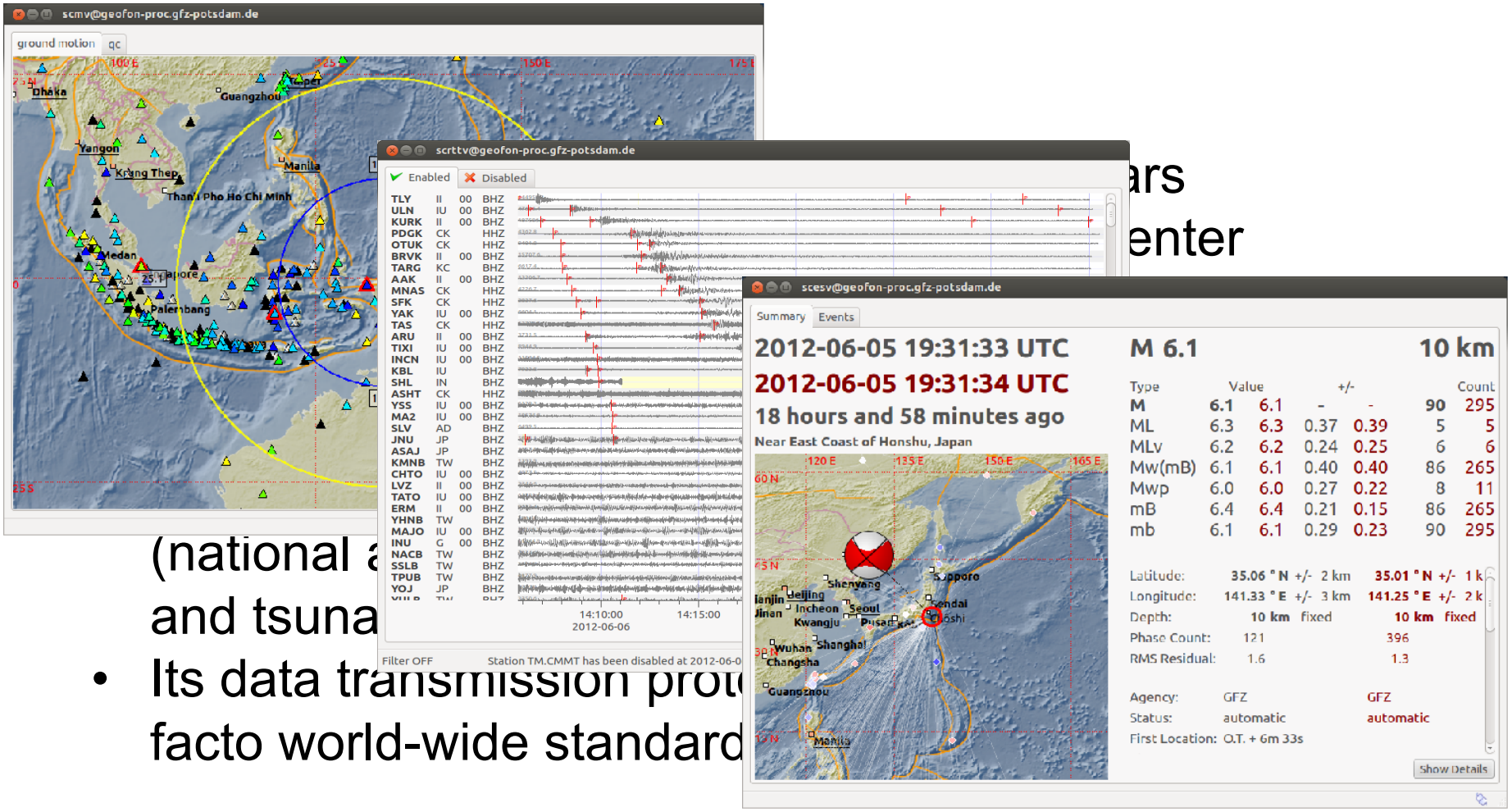
Two of the most popular GFZ web pages

~28.000 unique visitors / day (yearly average)

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

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

SeisComP3 software Architecture



ars

enter

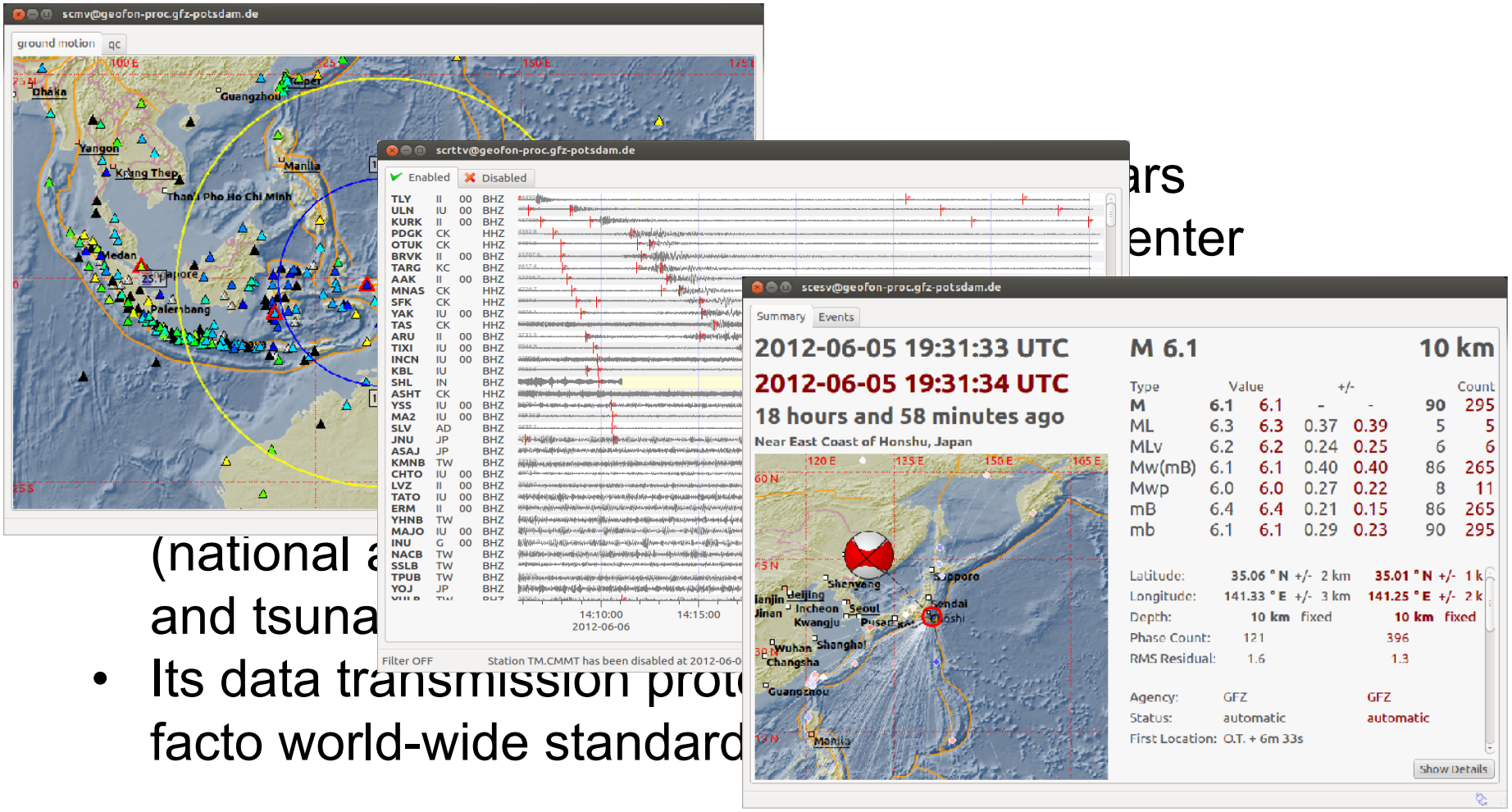
(national a

and tsuna

- Its data transmission prot

facto world-wide standard

SeisComP3 software Architecture



ars

enter

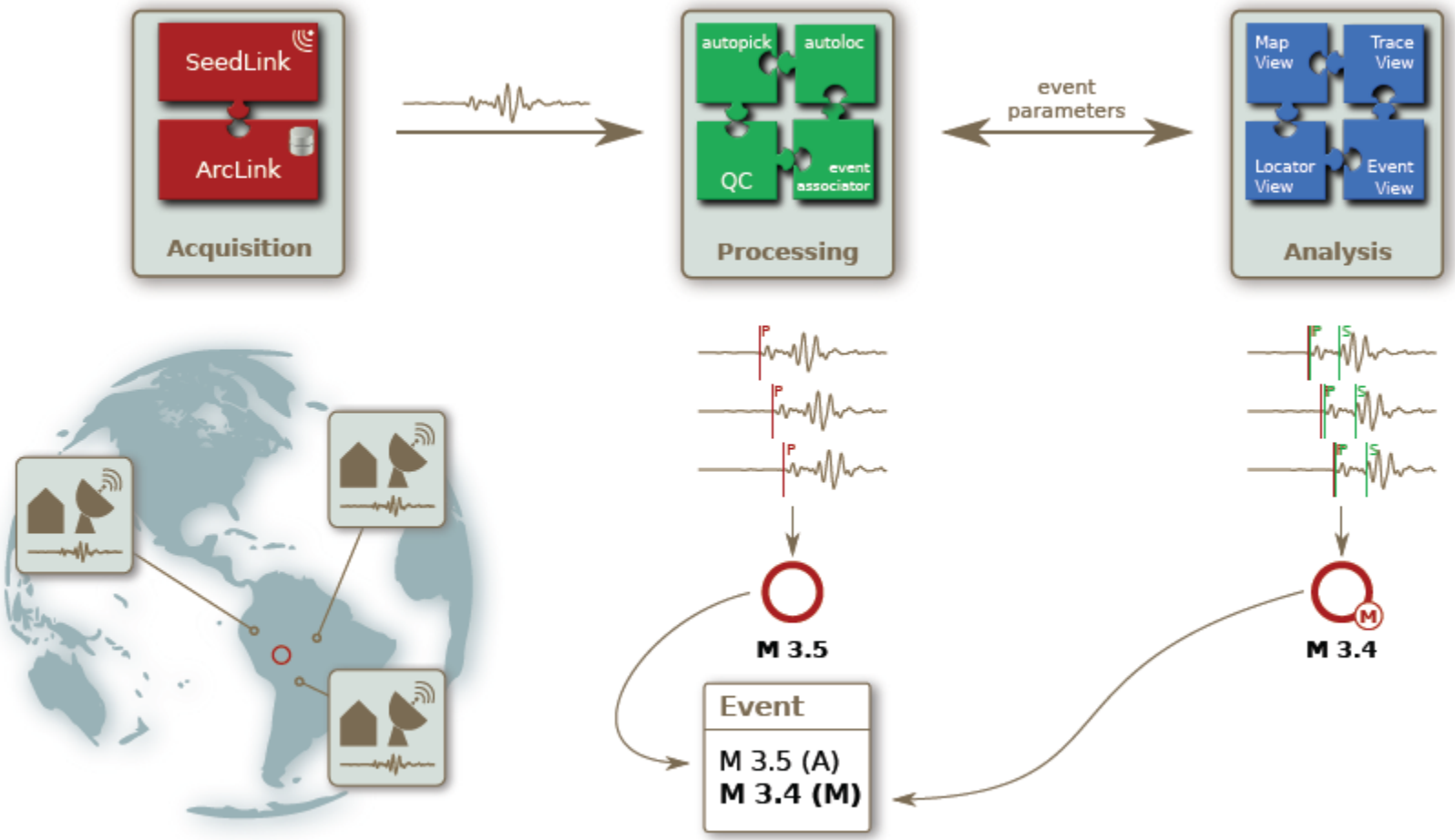
(national a

and tsuna

- Its data transmission prot

facto world-wide standard

SeisComP3 software Architecture



Courtesy of B. Weber and J. Becker

The 2014 Iquique Earthquake

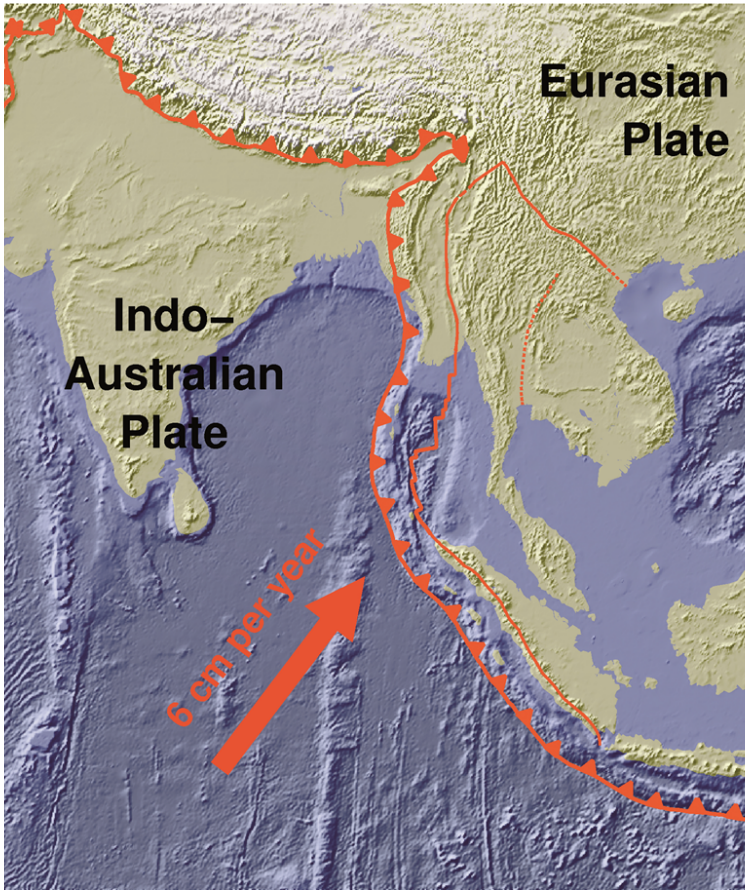
From rapid event information to scientific activities through HART deployment and data curation

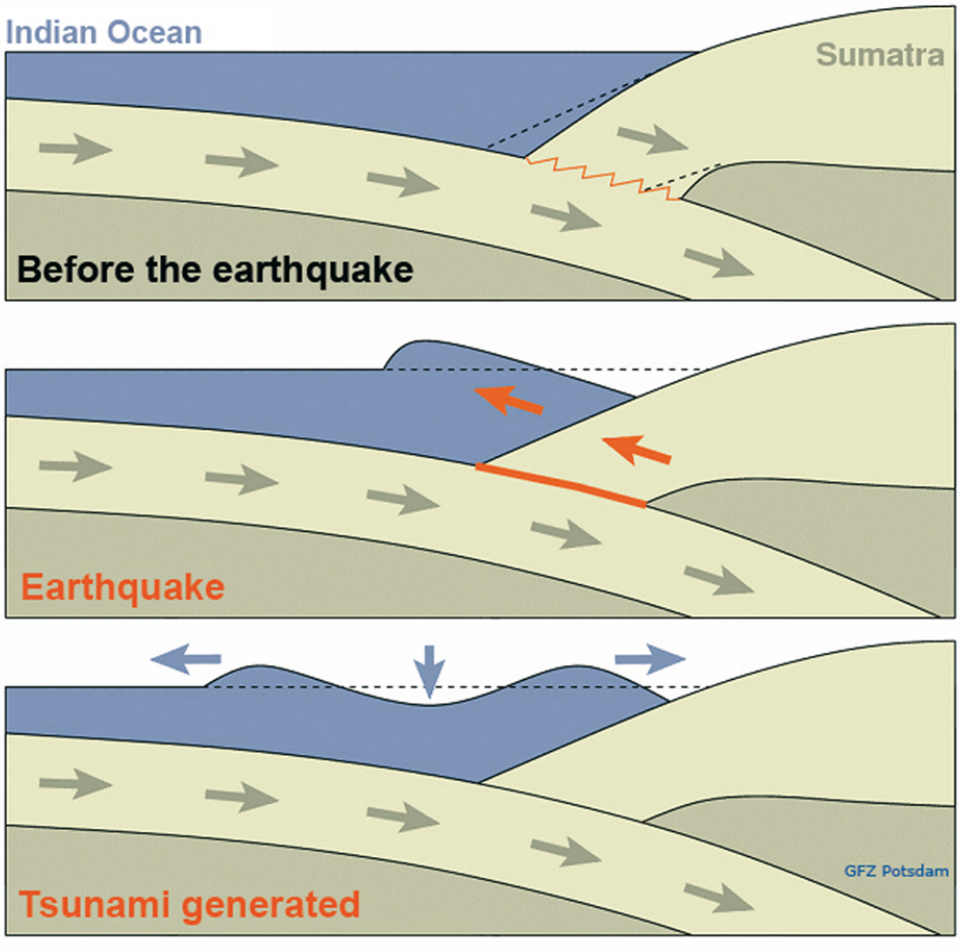
Time

- 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

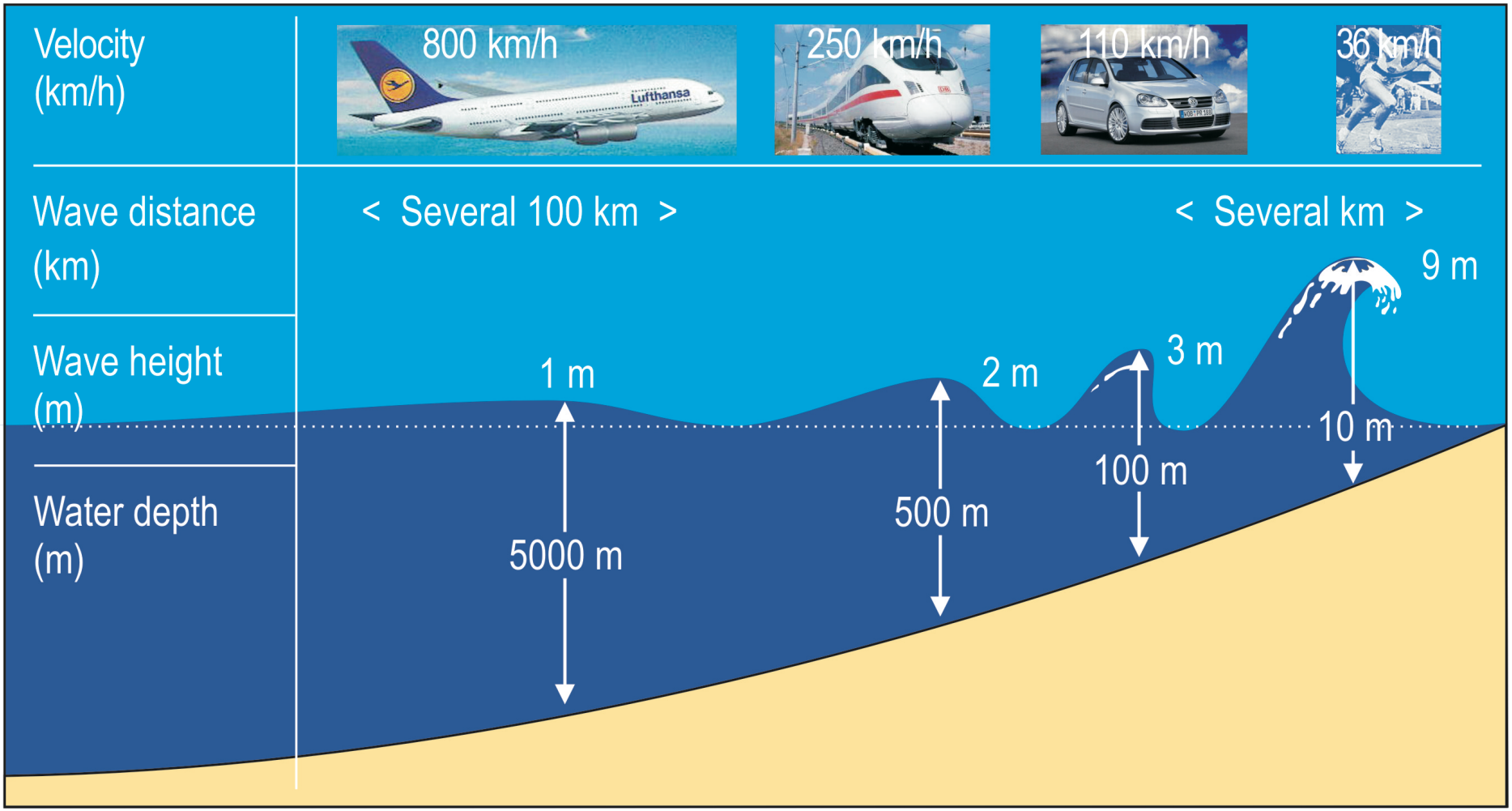
Tsunami Early Warning System

Tsunami generation



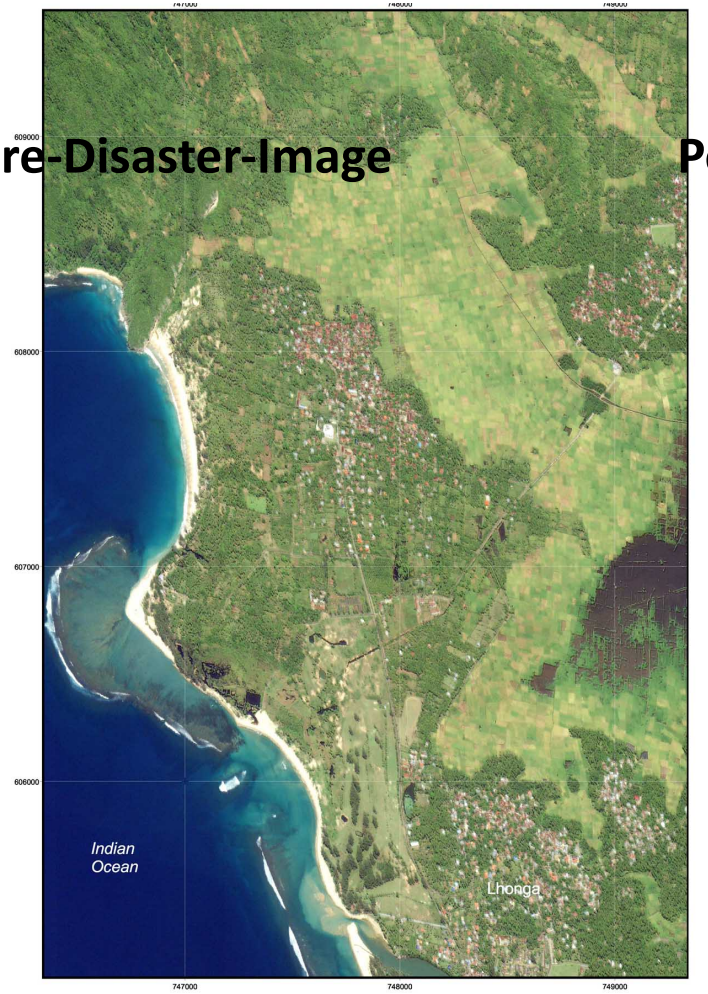


Tsunami propagation

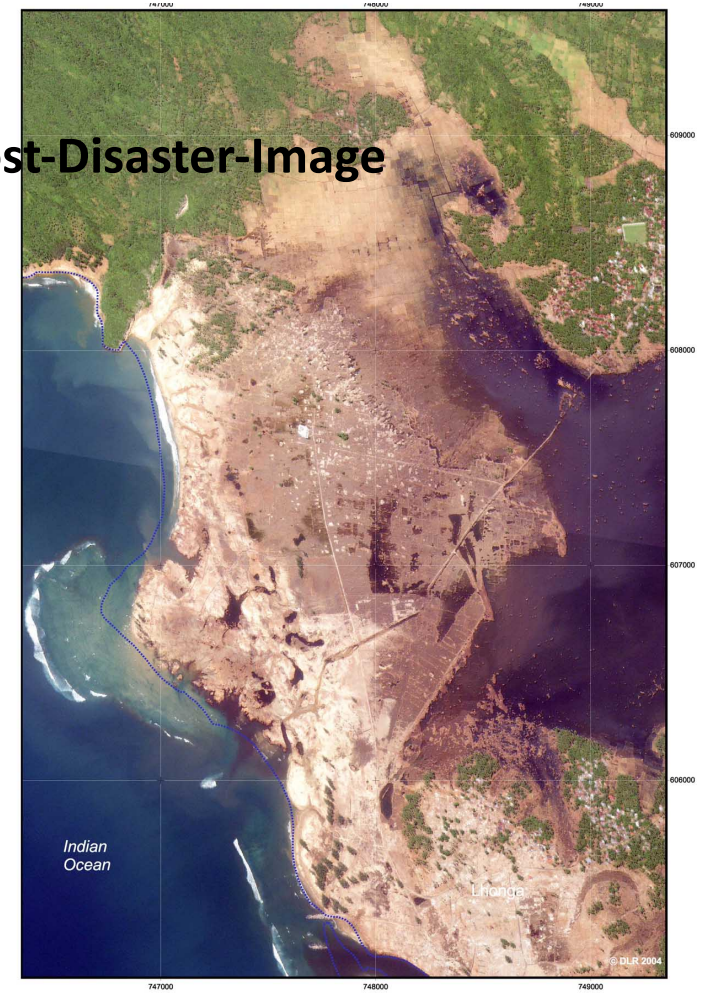


The 26th December 2004 - Banda Aceh

Pre-Disaster-Image

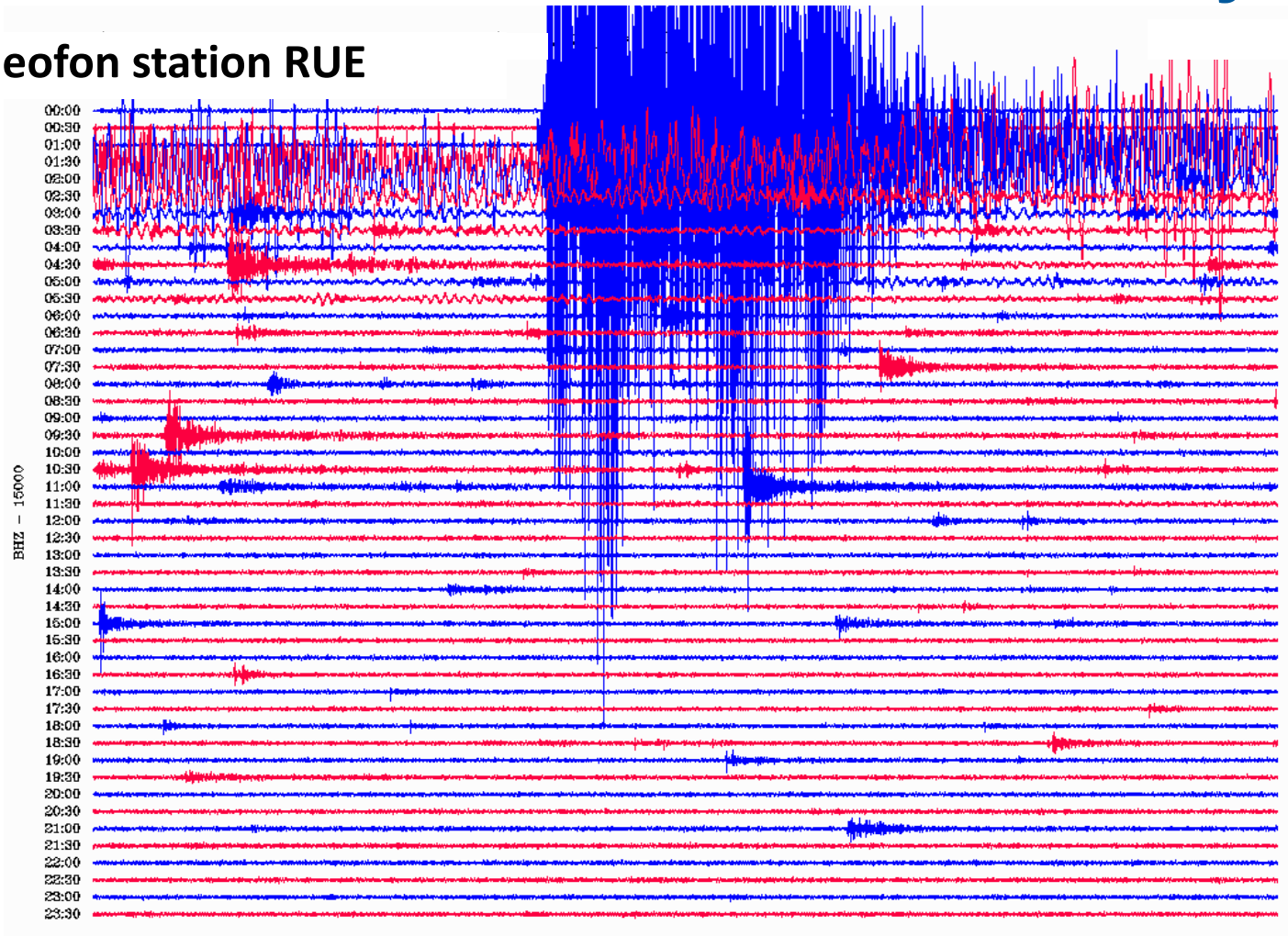


Post-Disaster-Image



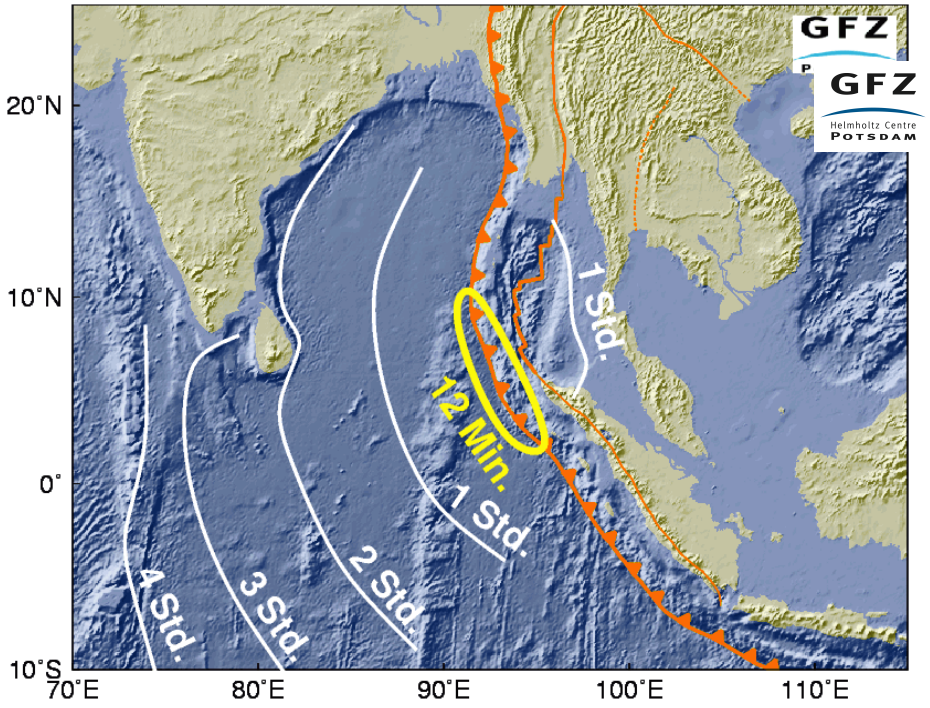
The 26th December 2004 - Germany

Geofon station RUE



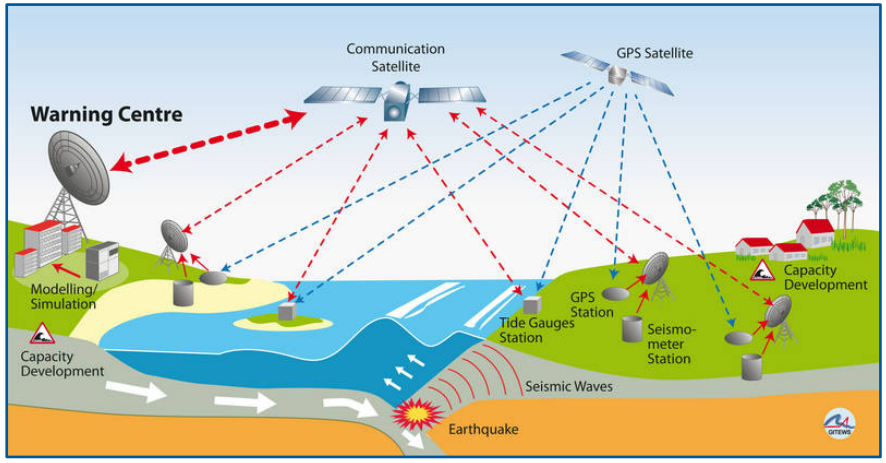
The Challenge, Tsunami Warning for Indonesia

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



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

The Challenge, Tsunami Warning for Indonesia



OT Earthquake signal recorded at the seismometer data transmitted to the DC
- → SeisComP3 automatic processing
- → SC3 operator starts manual waveforms picking

2 First/preliminary EQ Location (auto/man)

4 EQ parameters (Lat., Lon., Depth, O.T., Mag.) ready for dissemination. Decision making: disseminate?

5 → **DISSEMINATION PROCESS**

6 SC3 operator continues EQ analysis and updates the EQ parameters

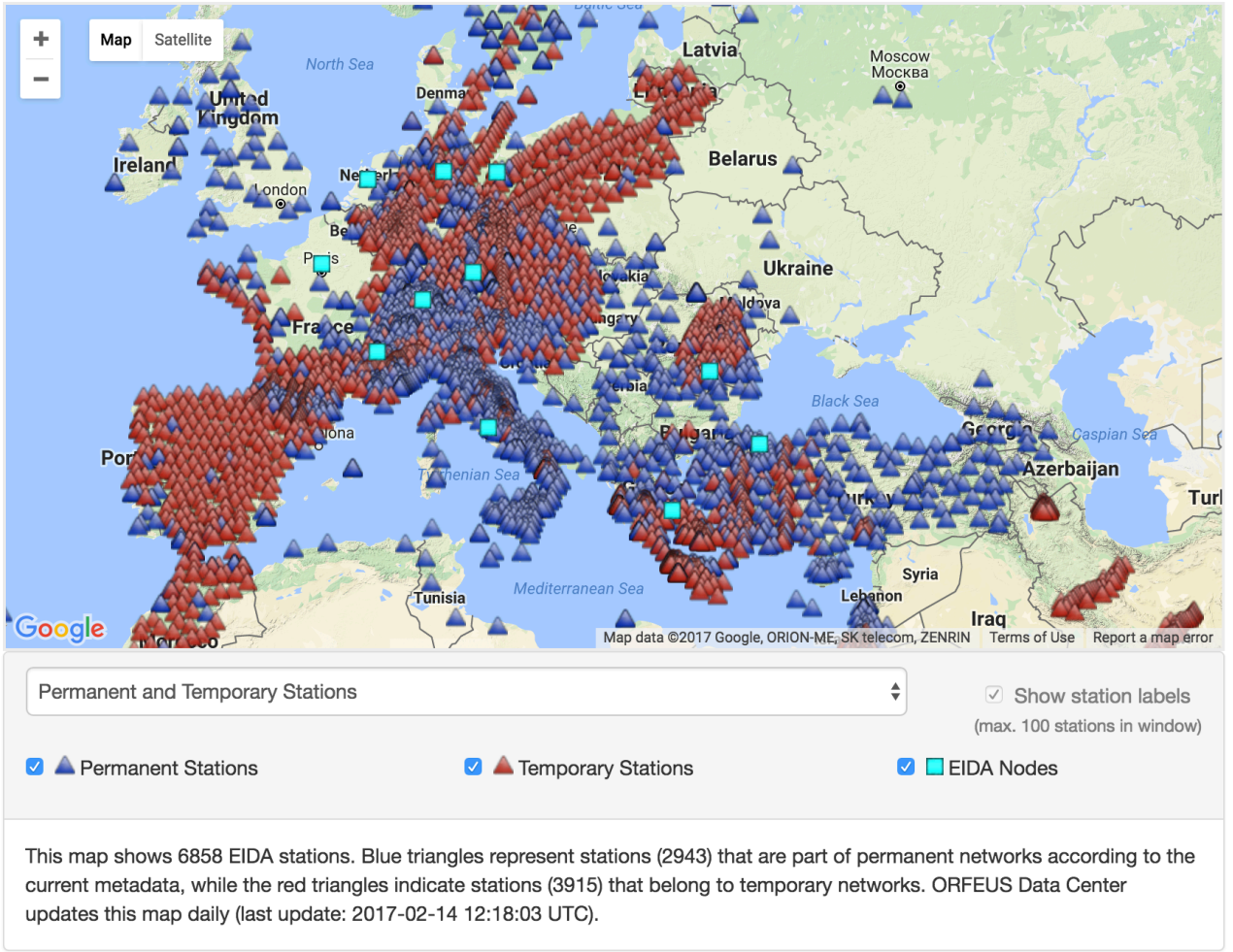
30min SC3 operator completed the EQ analysis and updates the EQ parameters (EQ param. review later…)

Monitoring room at Jakarta

From an isolated archive to an European Federation (EIDA/ORFEUS)

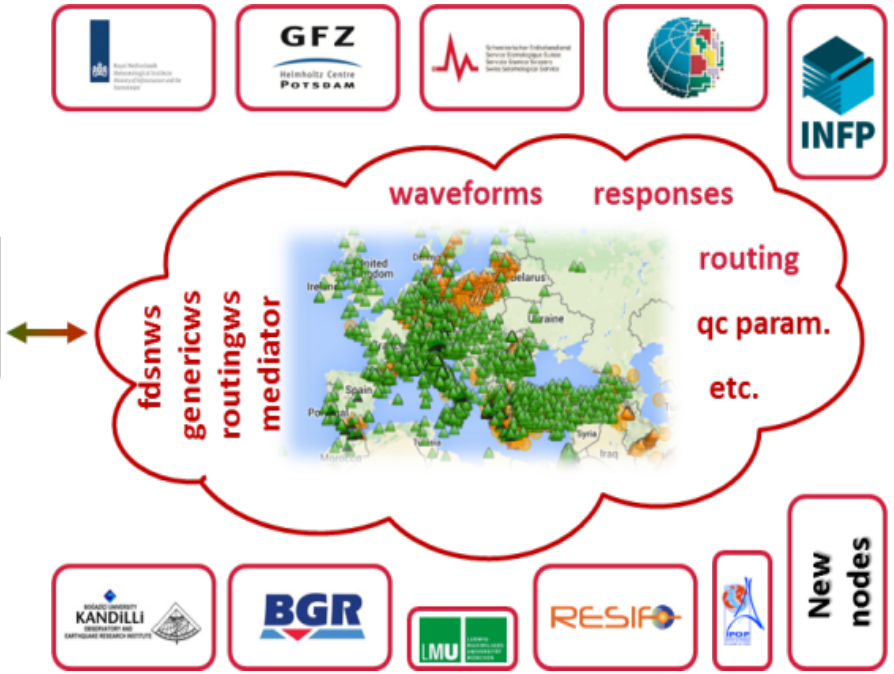
EIDA within ORFEUS

Safe, persistent archival and dissemination of seismic waveform data collected by European datacenters via distributed archives.



EIDA within ORFEUS

- Protocols to share data
  - Arclink protocol
- Synchronization of metadata
  - Daily
- Clients/tools development
  - SeisComP3
- Statistics
- Routing Table (maintenance)
- Common policies for f.i. data curation



Seismic Waveform Data

How to get waveforms and metadata: WebDC3 portal

http://eida.gfz-potsdam.de

FDSN Web Service Specifications

Version 1.1
2013/10/25

Purpose

To specify web service interfaces for the exchange of time series data, related metadata, event parameter and other data within the context of the International Federation of Digital Seismograph Networks (FDSN). The intention is to provide a specification that, when implemented at different FDSN data centers, can be used interchangeably by the same client software. The specification defines service names, query parameters and expected results.

List of service interfaces

dataselect – returns raw time series data in FDSN miniSEED format

station – returns metadata in FDSN StationXML and alternate formats

event – returns parametric data for events in QuakeML and alternate formats

Common service characteristics

Versioning

The services are versioned according the following three-digit (x.y.z) pattern:

SpecMajor.*SpecMinor*.*Implementation*

Where the fields have the following meaning:

SpecMajor: The major specification version, all implementations sharing this *SpecMajor* value will be backwards compatible with all prior releases. Values start at 1.

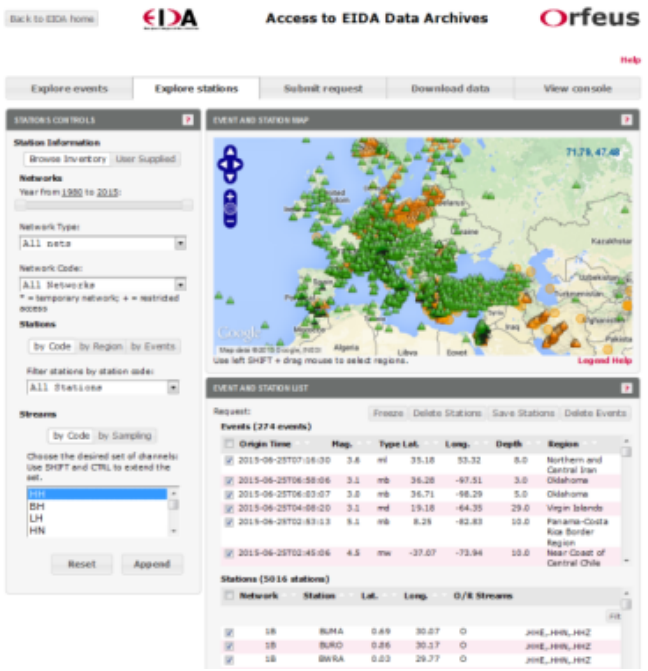
SpecMinor: The minor specification version, incremented when optional parameters or behavior is added to the previous specification but backwards compatibility is maintained with the previous major versions, i.e. all 1.# service versions will be compatible with version 1.0. Values start at 0.

Implementation: The implementation version, an integer identifier specific to the data center implementation. Useful to track service updates for bug fixes, etc. but with no implication on conformance to the specification.

Together the *SpecMajor* and *SpecMinor* versions imply a minimum expected behavior of a given service. This versioning scheme allows clients to expect specific behavior based on the *SpecMajor* version, while allowing the FDSN to extend the services with optional parameters while maintaining backwards compatibility. Each version number is service specific, there is no implication that *SpecMajor* version numbers across services are related.

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
- Help define seismological center best practices for ORFEUS community



EIDA NG: Why?

Challenges:

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

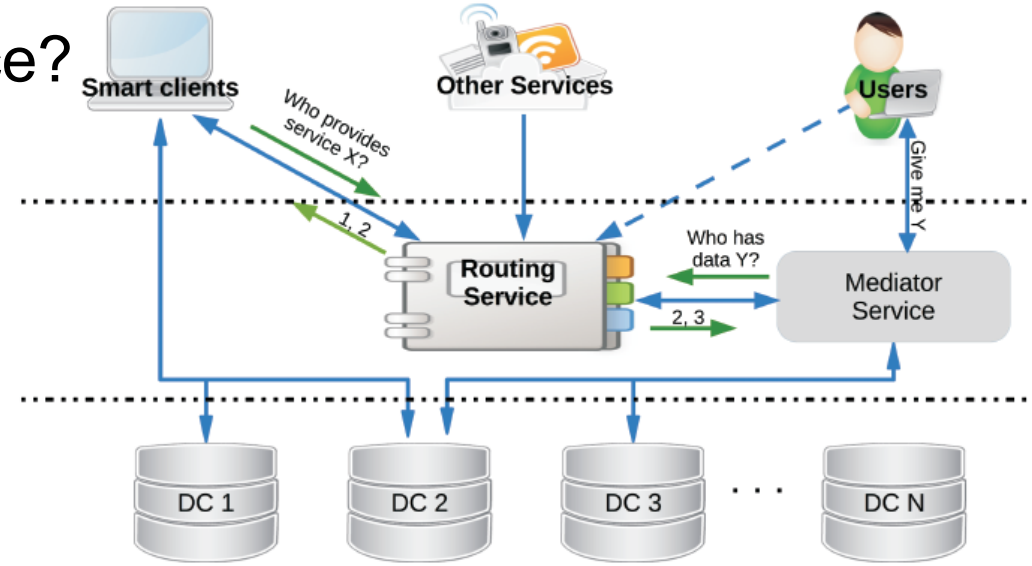
Requirements:

- 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
- Geolocation of data and services
- Contact information

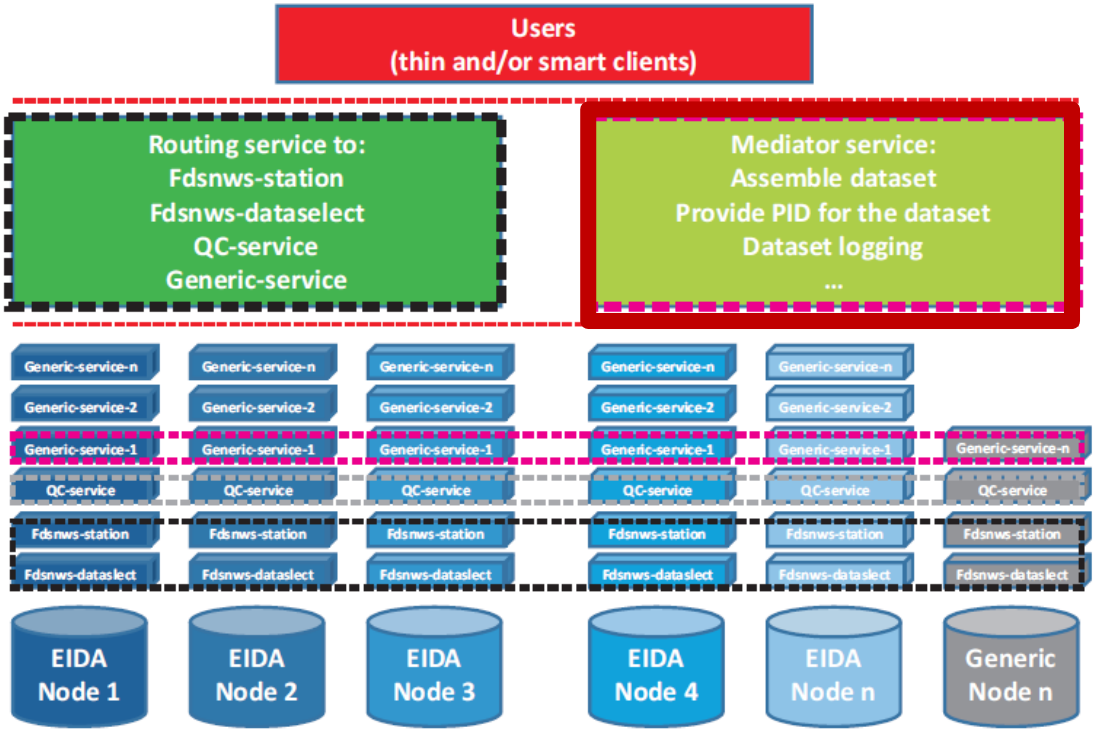


- Can also be deployed as a standalone router to run on the client side to create virtual DCs.

First stable release 10.2014

The EIDA Mediator (implementation 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.

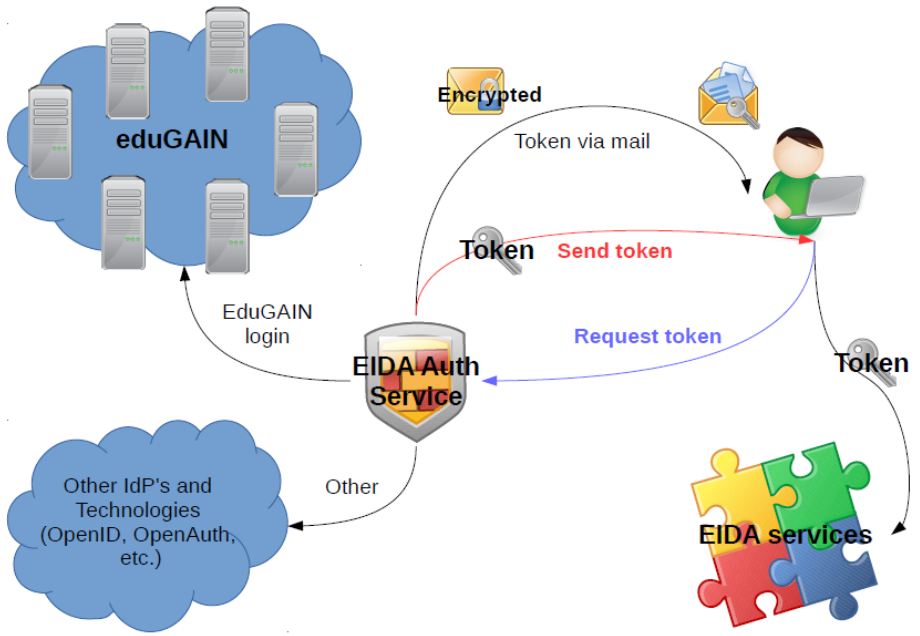


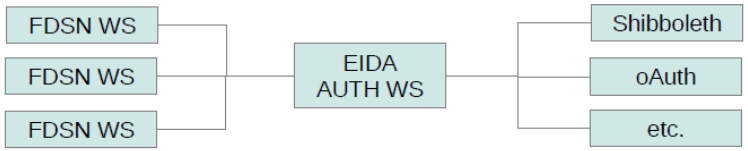
Latest developments and improvements EUDAT2020 and EPOS

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).





How to cite a dataset or a data provider?

- Although this is recognized by the scientific community, there was no standard way to cite or acknowledge them.
- Often, distributing data centres are credited in acknowledgments with no mention of the data generators. DOI metadata allows recognition.
- Acknowledgments can be inconsistent; acknowledgement sections are not systematically indexed and therefore it is hard to quantify the scientific “impact” of a seismic network. DOI is a simple, unique string.
- Data is distinct from the scientific ideas presented in a paper. Citing a scientific paper in an attempt to acknowledge data generators, is unsatisfactory.

DOIs for Seismic Networks

The EIDA Data Centers in Europe and the IRIS DMC in the US initiated a joint effort to define the usage of DOIs for seismic networks. The FDSN through its WG III on Products, Tools, and Services has agreed upon a method to provide attribution to permanent and temporary seismic networks using FDSN Network Codes. This has been reviewed by FDSN WG III and has been agreed to by consensus.

http://www.fdsn.org/wgIII/V1.0-21Jul2014-DOIFDSN.pdf

Evans, et al (2015)

Do you want me to archive your data?

Data management agreement

We thank you for entrusting us with your network's data, and for your use of our services.

We offer you long term preservation of your data in the same conditions as we keep our own (see details below). However, we cannot provide any formal guarantee for its preservation.

We would like to make you aware of a few points regarding our data management policies.

1. Data surety

We will store two copies of your data in different locations at our campus.

We may hold a third copy off site if needed.

2. Your role as Principal Investigator (PI)

You are the PI for this network, and we will contact you whenever issues arise in relation to your data.

Do you want me to archive your data?

Data management agreement

We thank you for entrusting us with your network's data, and for your use of our services.

We offer you long term preservation of your data in the same conditions as we keep our own (see details below). However, we cannot provide any formal guarantee for its preservation.

We would like to make you aware of a few points regarding our data management policies.

1. Data surety

We will store two copies of your data in different locations at our campus.

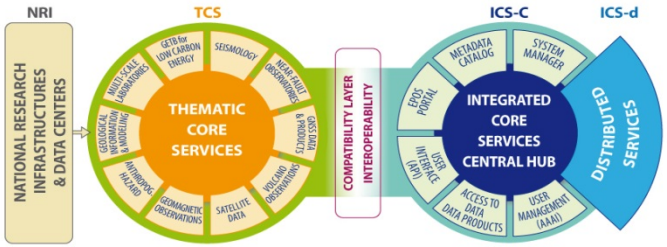
We may hold a third copy off site if needed.

2. Your role as Principal Investigator (PI)

You are the PI for this network, and we will contact you whenever issues arise in relation to your data.

EIDA/ORFEUS within EPOS

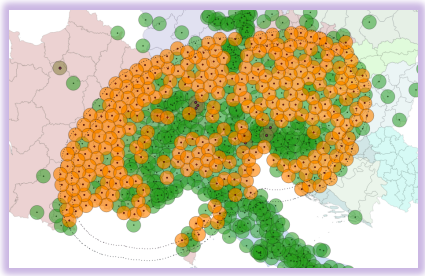
Long term plan to create a single, pan-European, sustainable and distributed infrastructure for solid Earth science to support a safe and sustainable society.



- EIDA is a core component of EPOS Seismology;
- EIDA is the prototype service to define and demonstrate compliance with the emerging EPOS ICS (Integrated Core Services).
- EIDA and other EPOS TCS (Thematic Core Services):
  - EIDA provides a template governance and technical architecture that may be followed by other TCS.

EIDA/ORFEUS within EPOS

Long term plan to create a single, pan-European, sustainable and distributed infrastructure for solid Earth science to support a safe and sustainable society.



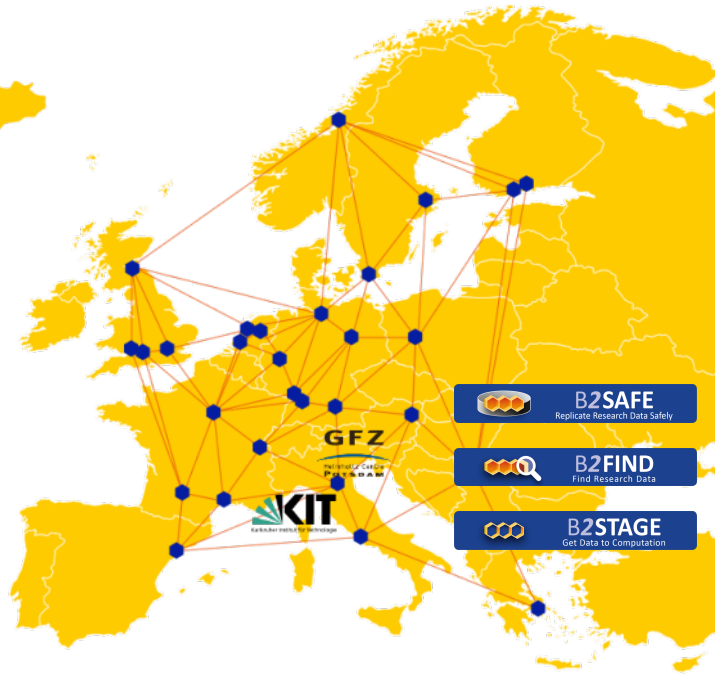
- Acceleration and Strong-Motion data
- Volcanology: seismic and infrasound monitoring
- Ocean Bottom Seismometers
- Structural monitoring
- Near-fault observatories

EUDAT2020 - 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 (as EPOS member):

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

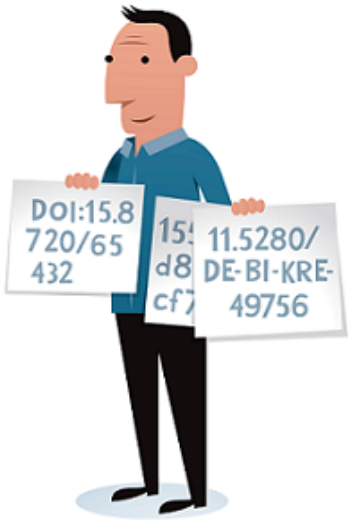


EUDAT

Persistent Identifiers and attribution

DOI for Networks

- One of the main driving forces for the adoption of DOIs for seismic networks.
- First step to the proper attribution of data usage in scientific articles.
- Now DOIs appear in community web pages.



PID for data file

- Persistent ID for data files (~25 millions)
- Facilitate data discovery and its staging directly into the computational facility
- Include metadata related to f.i. curation, and versioning.

Data Collections

Why do we need it?

- We cannot afford to store all datasets (~5 million/year).
- Reproducibility: same request more than once can result in different datasets.
- Provide immutable network/event based datasets.
- Versions/Dynamic data.

How can we benefit from PIDs?

- Datasets defined as a collection of PIDs.
- Metadata associated with the data collection.
- Accurate statistics by monitoring of PID resolution.
- Minimize the data movement.
- Fast transmission of data collection.
- Facilitate the staging on computer facilities.

“Old” Challenges in seismology

- Standard Waveform Data Format
- Merging of different data sets
- Standard (Real-Time) Data Exchange Protocol
- Virtualization of Seismological Networks
- Standard Data Archive Access Protocol
- Virtualization of Seismological Data Archives
- Synchronization of distributed data sets
- Combining data and HPC services

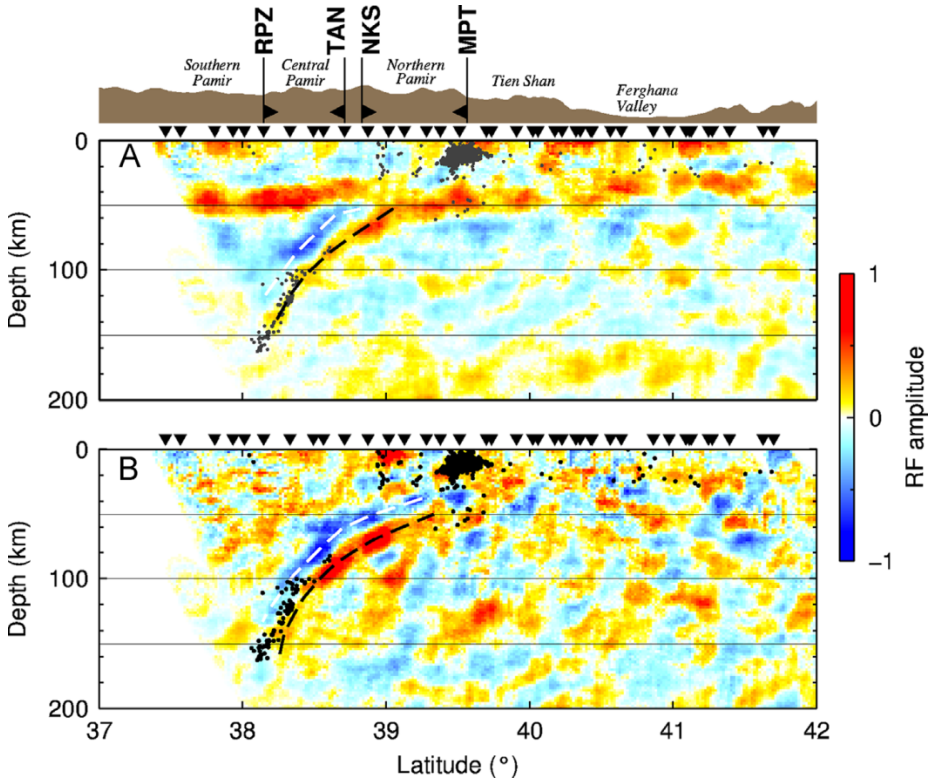
Challenges in the seismological community

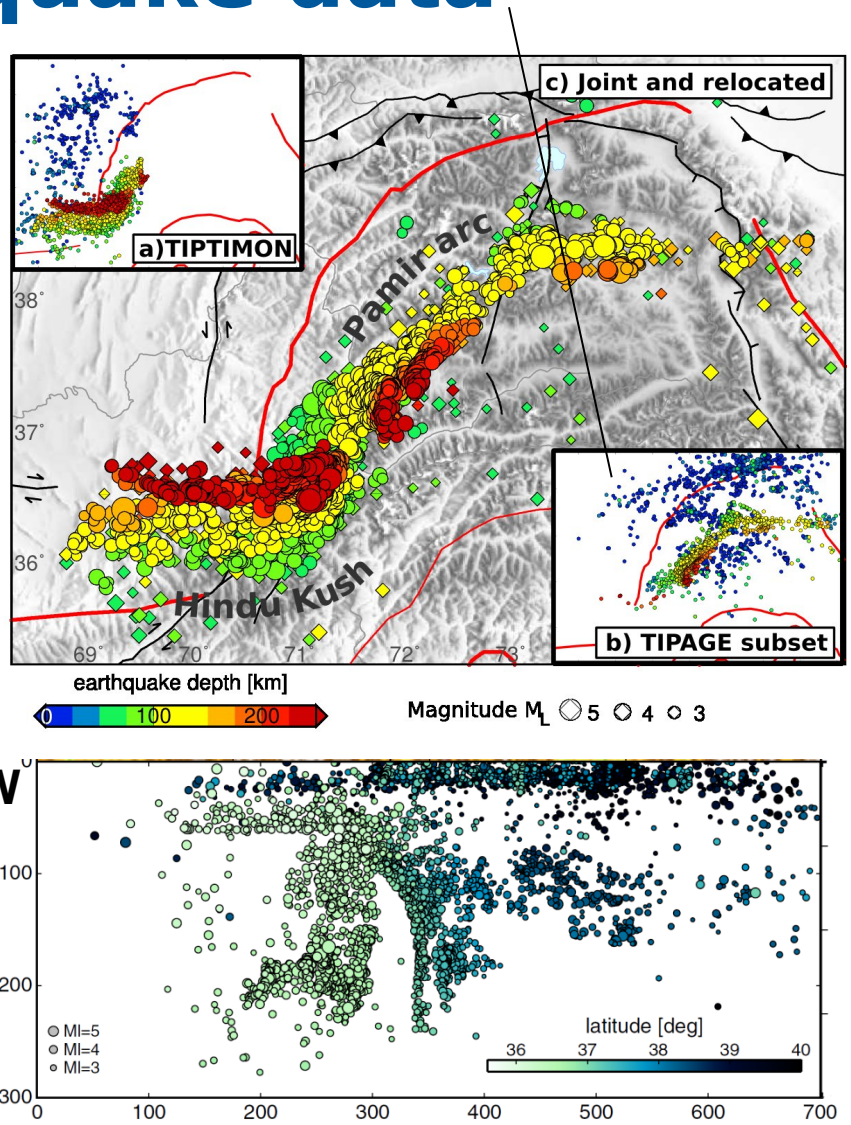
- Dynamic Data
  - ✓ Waveforms can be modified during the first days.
- Reproducibility of experiments
  - ✓ Freezing datasets in time
  - ✓ Millions of requests/year referencing thousands of records
  - ✓ Before and after curation of data/metadata
  - ✓ Real-time data seems to be much more complicated for the time being.
- Integration with other Earth sciences
  - ✓ Not enough standards in other branches of science
  - ✓ Different levels of evolution regarding infrastructure
  - ✓ Protocols and data formats
  - ✓ Semantics! What does f.i. “deptht” means?

Thanks for your attention!

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)