

GFZ

Helmholtz-Zentrum
POTSDAM



Javier Quinteros, GEOFON Data Centre (Deutsches GFZ)

Yerevan, September 25th, 2019

Research Fields of GFZ

Georesources

**Modular Earth Science
Infrastructure (MESI)**

**Planet Earth
Monitoring**

**Natural Disasters
& Risk Reduction
Strategies**

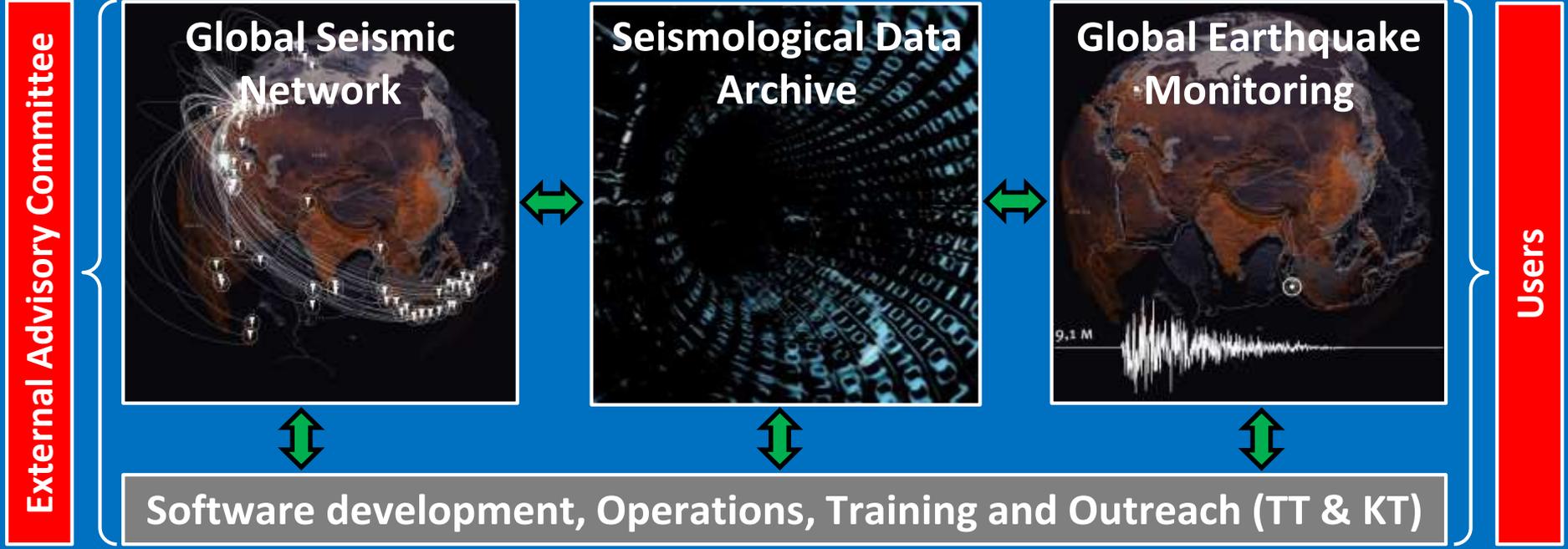
**Earth Surface and
Climate Dynamics**

**Continental
Dynamics**

The GEOFON Program

User
Facilities
MESI

OPEN
DATA



People



A. Strollo



T. Zieke



L. Hillmann



A. Heinloo



J. Saul



S. Hemmleb



J. Quinteros



R. Zaccarelli



P. Evans



F. Tilmann

GFZ

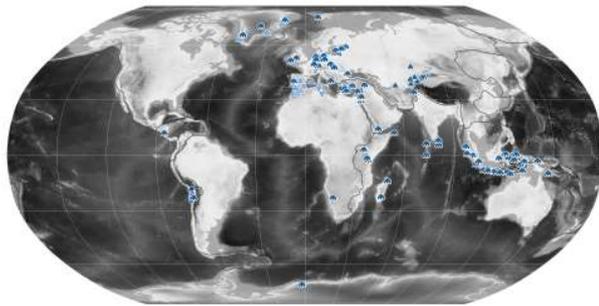
Helmholtz Centre
POTSDAM

HELMHOLTZ
ASSOCIATION

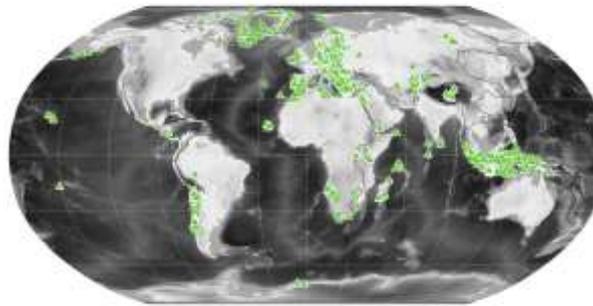
The GEOFON Program

GEOFON in a nutshell:

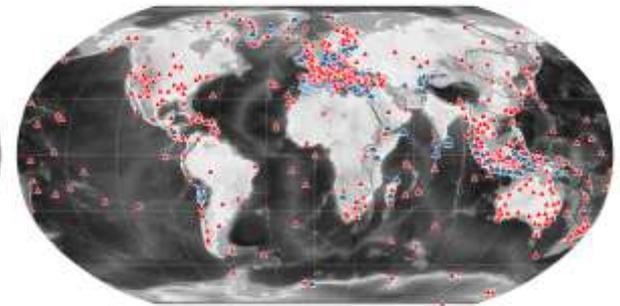
- A global real-time high quality seismic network
- The largest seismological data archive in Europe
- The fastest earthquake information systems worldwide
- A competence center for IT developments in seismology



**100+ GEOFON
real-time stations**

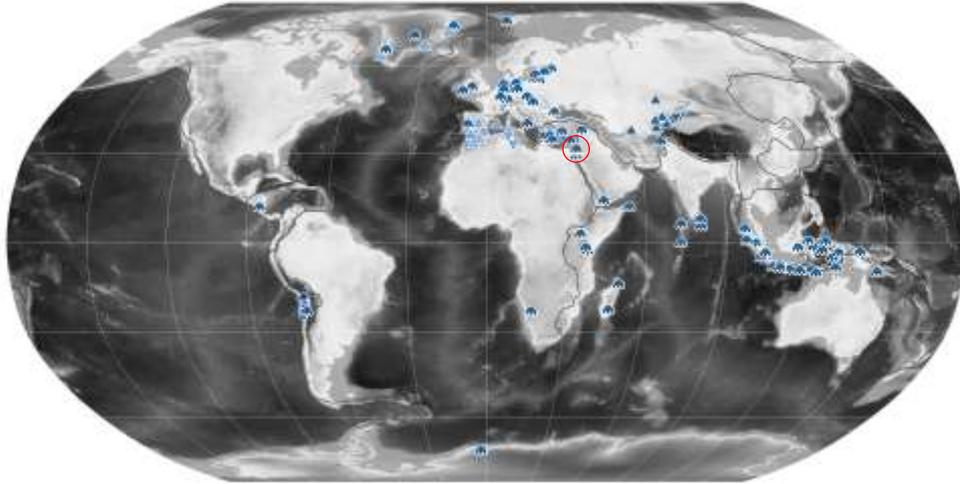


**Archive data holdings
3000+ stations
Dense regional networks
GFZ/GIPP seismology
experiments**



**Extended Virtual real-time
Network 1000+ stations**

The GEOFON Global Seismic Network



Scope

- Provide high quality real-time data for rapid and global earthquake monitoring
- Foster international cooperation

Facts

- 84 GE stations
- 40+ affiliated stations
- Open real-time data distributed to 300+ data centers worldwide
- 6500+ events and 1500+ EQ mechanisms (MT) published per year
- 200.000+ notifications/year to various stakeholders via e-mail and SMS

GEOFON Data Centre (1993): GEOFON Seismic Network. Deutsches GeoForschungsZentrum GFZ. Other/Seismic Network. doi:10.14470/TR560404.

The GEOFON Global Seismic Network

How to join and start sharing seismic stations and data?

- Sent a proposal to GEOFON, if the proposed cooperation fits the GEOFON mandate a MoU may be negotiated and signed.
- Within the bilateral MoU a general frame of cooperation will be defined and the contribution to the cooperation from both sides will be described.

Memorandum of Understanding for Seismological Coordination, Cooperation and Collaboration among the institute name (acronym) and the German Research Centre for Geosciences (GFZ)

Whereas the acronym and the GFZ will operate jointly a broadband seismograph station for global and regional studies of the earth, and

Whereas the acronym and the GFZ are committed to making the data from this seismographic station openly available to the global seismological community, and

Whereas the acronym and the GFZ wish to underscore the importance of international cooperation and their commitment to the goals of the Federation of Digital Seismographic Networks;

The acronym and the GFZ will operate jointly the seismic station in the Salta province, Argentina associated to the GEOFON network. Data from the station will be shared under the principle of free and equal access. Additional sites may be added under this agreement by the consent of representatives of both parties.

Understandings:

- The acronym and the GFZ will coordinate and share long-term maintenance responsibilities.
- The GFZ will make facilities available for training station field maintenance for technicians from the acronym.
- The GFZ and the acronym will cooperate in data processing and provide for equal and open access to all data of the GEOFON network or broadband stations operated by the acronym.
- The GFZ and the acronym will designate representatives to implement this Memorandum of Understanding through an implementation plan.

Implementation Plan Elements

Site

The acronym and the GFZ have selected a site in the xxx close to the town xxx (150 km from the city of xxx). The goal for the site preparation is the construction of the instrumental bunker. The base for mounting of STS-2 sensor should be stable and the bunker secured to avoid vandalism.

Instrumentation

GFZ will provide the following station equipment:

Quaterna Q330 data logger
Stevenson STS-2 broadband seismometer
a strong motion sensor
a SeisComP data recorder and communication unit, a power back up unit and auxiliary equipment.

If the data is first transferred from the station to the acronym and then to the GEOFON, the acronym should operate a PC with SeisComP software. This PC will be in this case provided by the GFZ.

Maintenance

Maintenance at the station will be primarily the responsibility of the acronym. The GFZ will provide training for their technicians if necessary. The training will be done primarily during the installation or later on site. The acronym will provide for the basic travel expenses (transportation) for any additional training in Potsdam. The GFZ will then provide the incidental costs of the stay in Potsdam. GFZ will provide reasonable means (GGK, insurance) for vehicle damaged or broken equipment and any additional expenses, which might be needed. This includes exceptional site visits.

Data

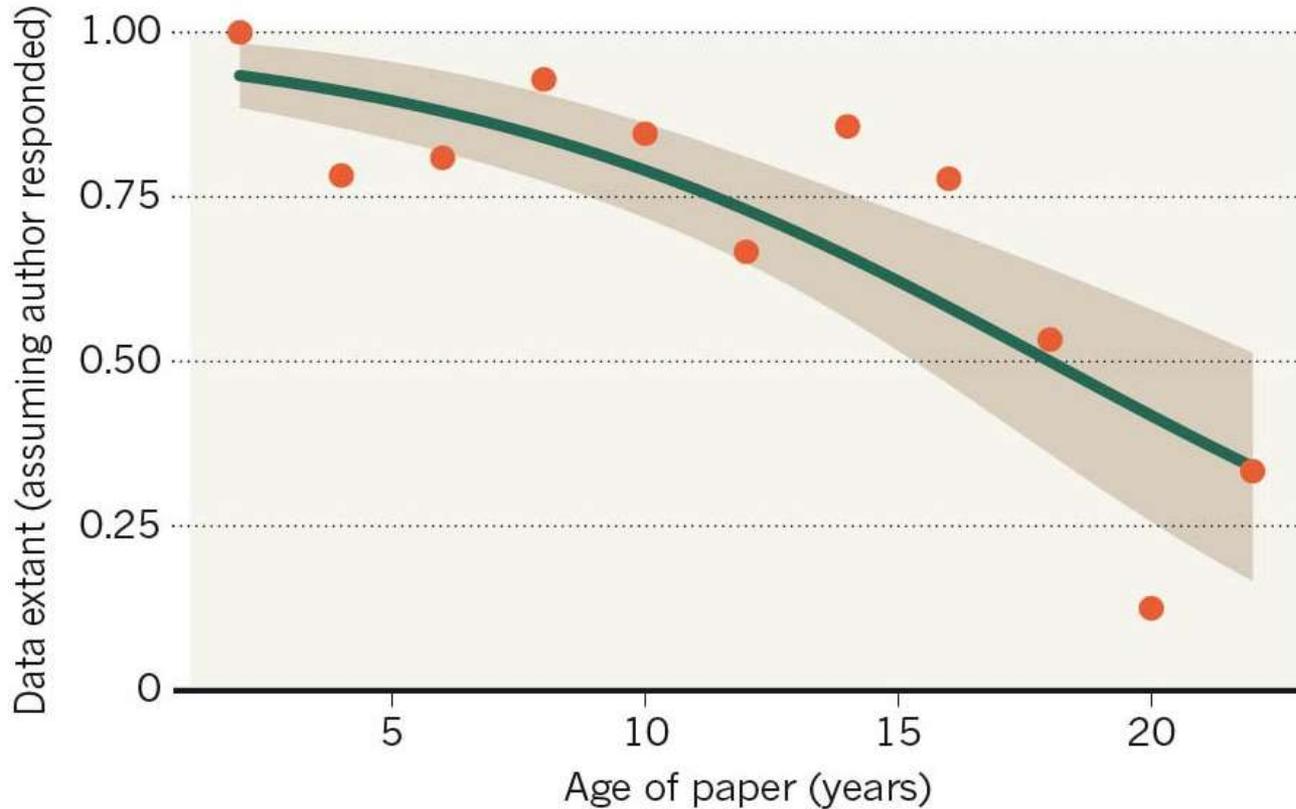
The station will be an open station. Internet access to the station will be provided by the acronym. The station should be accessible through telephone lines. GFZ Internet access (mobile means (GGK, insurance) for remote control. If a connection is available, the access should be directly granted to the GEOFON by the acronym independently to avoid overlapping of tasks from both sides. They should have direct access to the IP address of the station. The data will be transferred via Internet to acronym and GFZ and will be made available for immediate public access at GFZ Seismological Data Center.

GFZ will make its data processing software available to the acronym and vice versa.

<http://geofon.gfz-potsdam.de/>
geofon@gfz-potsdam.de

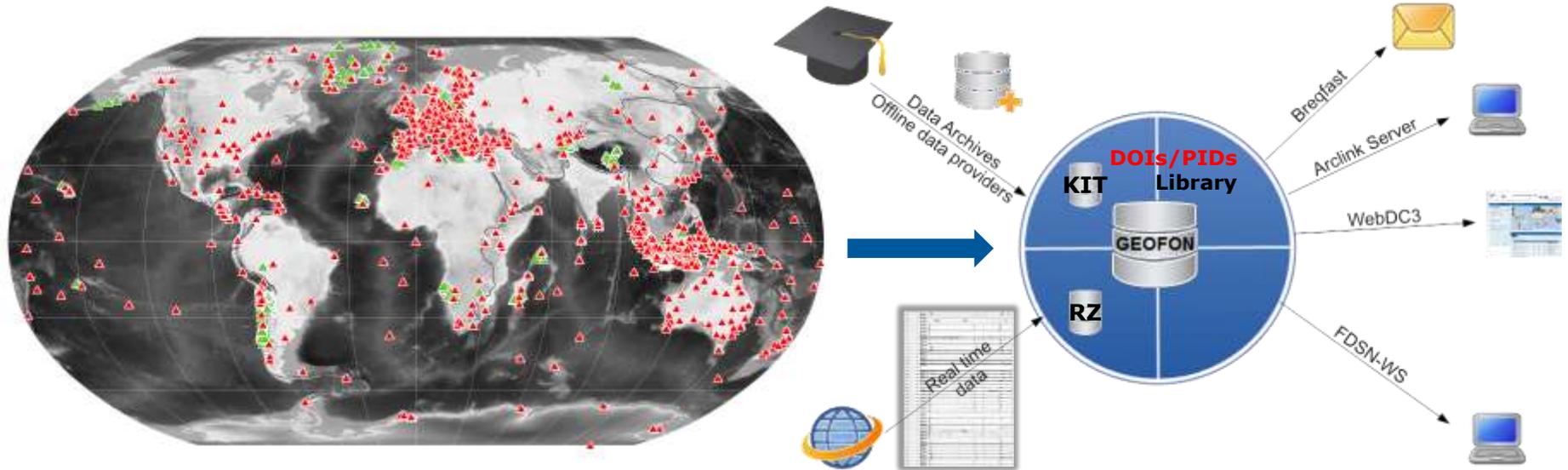
What happens to data after publication?

As research articles age, the odds of their raw data being extant drop dramatically.



Vines, T. H., et al. (2013) doi:10.1016/j.cub.2013.11.014

The GFZ seismological data archive



Scope

- Facilitate data access and usage for scientific research
- Promote in-house software development

Facts

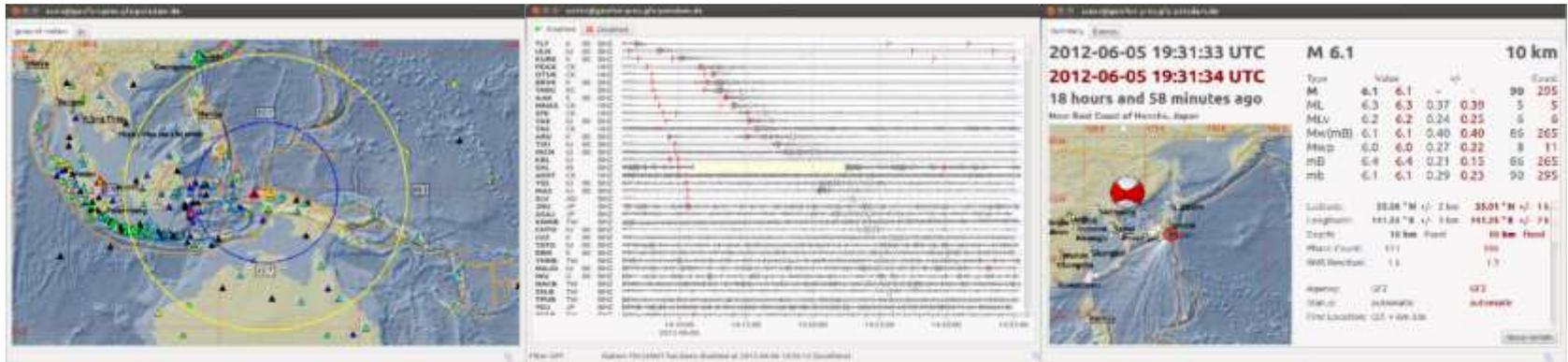
- 100 TB, 3600+ stat.
- 20 GB/day ingested
- 270 GB/day distributed, mostly in real-time

Challenges

- Continue driving development within the European Integrated Data Archive (EIDA)
- Improve attribution and discovery
- FAIR data

Software developments: SeisComP3

- In-house development over the last 10 years
- Base for operation of the GEOFON data center (archive and realtime) and rapid earthquake information
- Widely used package for seismological data acquisition, data processing and data exchange (national and international earthquake observatories and tsunami warning centres)
- Its data transmission protocol SeedLink became a de facto world-wide standard for realtime transmission



- SeisComP3 is on GitHub at <https://github.com/SeisComP3> and used worldwide (400+ free non-commercial licenses)

<http://www.seiscomp3.org/>



Big Data Analytics

Big Data:
Data sets that can hardly be processed and analyzed by traditional methods because of their volume, complexity, or structure

Value of the data
Selection of relevant data

Reliability of data
e.g., uncertainty in proxies or ensembles

Large data sets
e.g., time series, simulation output

Fast process
e.g., real-time data

Volume

Value

Big Data

Velocity

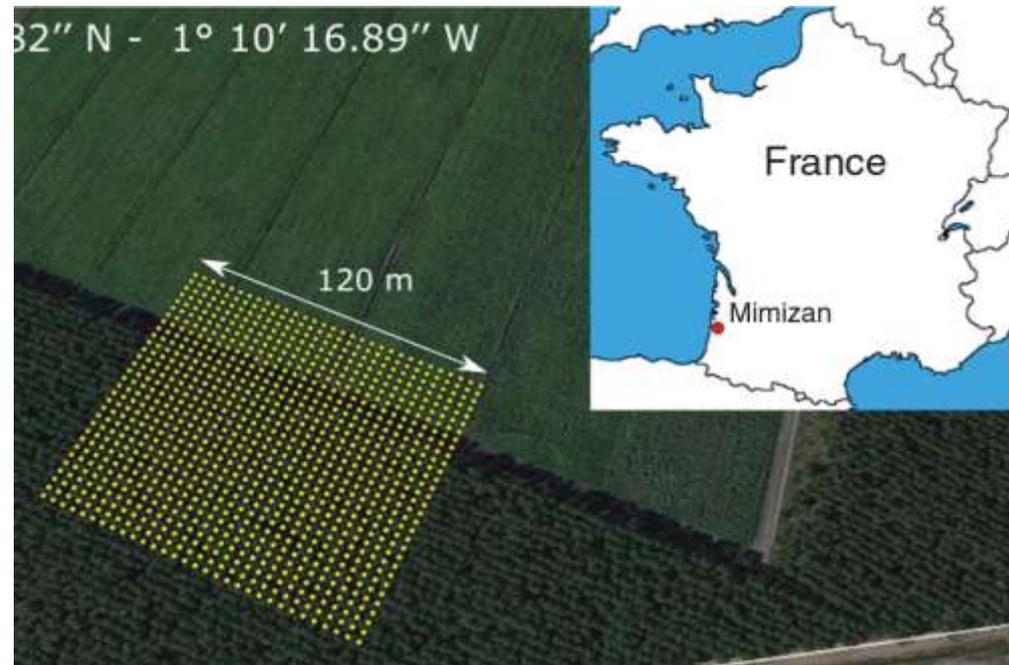
Veracity

Variety

Different data sets
e.g. observation, simulation, historical, citizen science

Large-N experiments and new technologies

- The demand for new dense observations is advancing fast.
- Possible adoption of different data formats to archive and provide data.
- Evaluation data formats to store big volumes of wave-forms.
- Ingestion of data to HPC facilities.
- Extension of the Dataselect-WS to support the creation of standard miniSEED data in runtime and different on-the-fly conversions.

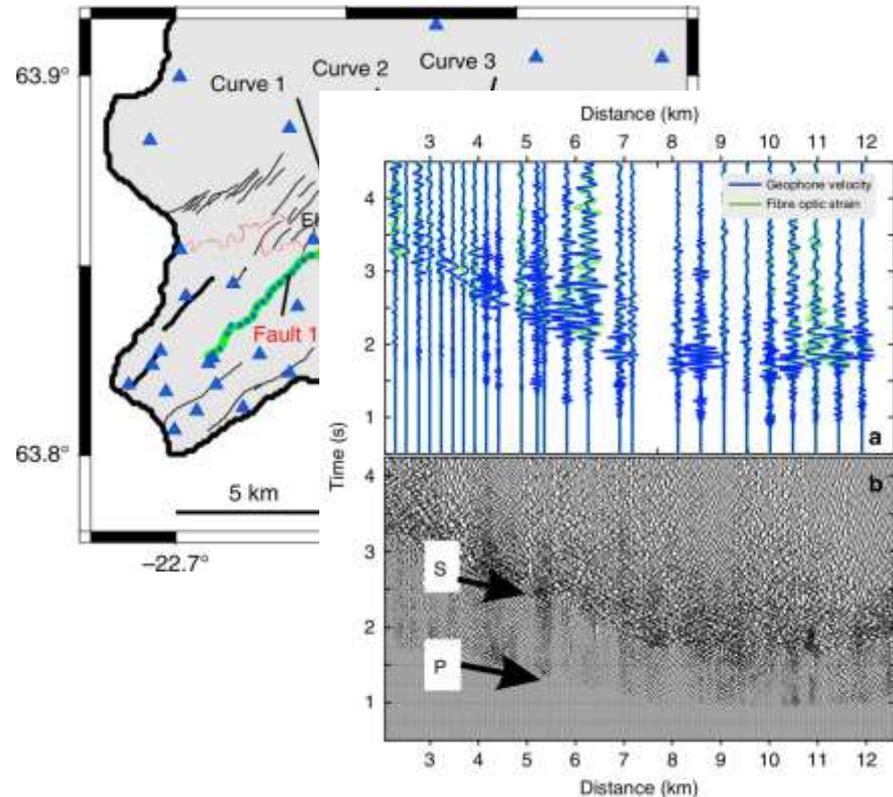


Roux et al., SRL (2018) doi: 10.1785/0220170196

Long, dense and... big data

Surface cable: distributed acoustic sensing (DAS) data characteristics:

- 15 km long
- One trace every 4 m
- Acquisition rate 1000 Hz
- 3700 traces
- 1 Gb/mn
- 12 Tb/week

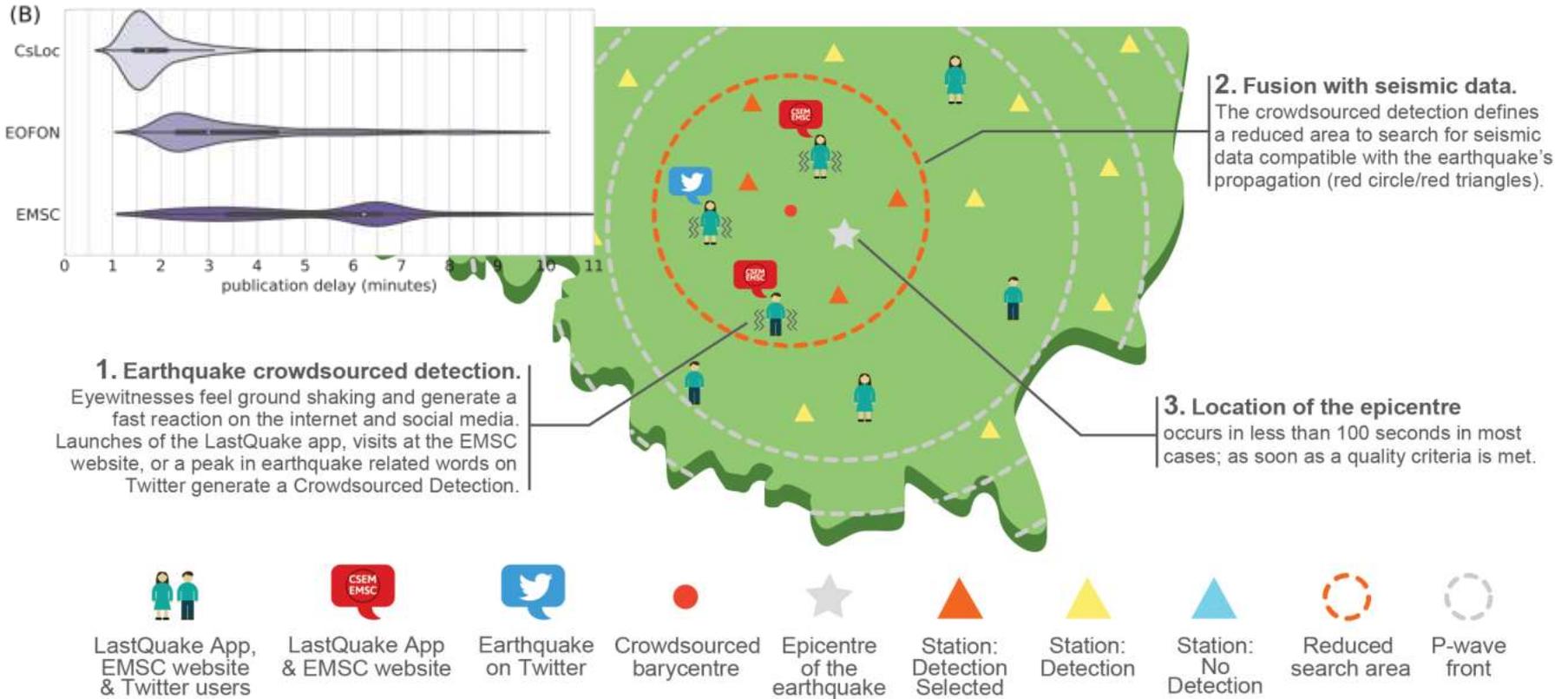


- *Etna volcano 2018*
- 4 km long, spatial sampling 2 m, 1000 Hz, 20 days: 8.3 Tb
- 1.3 km long, spatial sampling 2 m, 1000 Hz, 15 days: <1 Tb
- 21 km long, spatial sampling 2 m, 1000 Hz, 1 week: >11 Tb

Jousset, et al. (2018) doi:10.1038/s41467-018-04860-y

CrowdSeeded Locations (CsLoc)

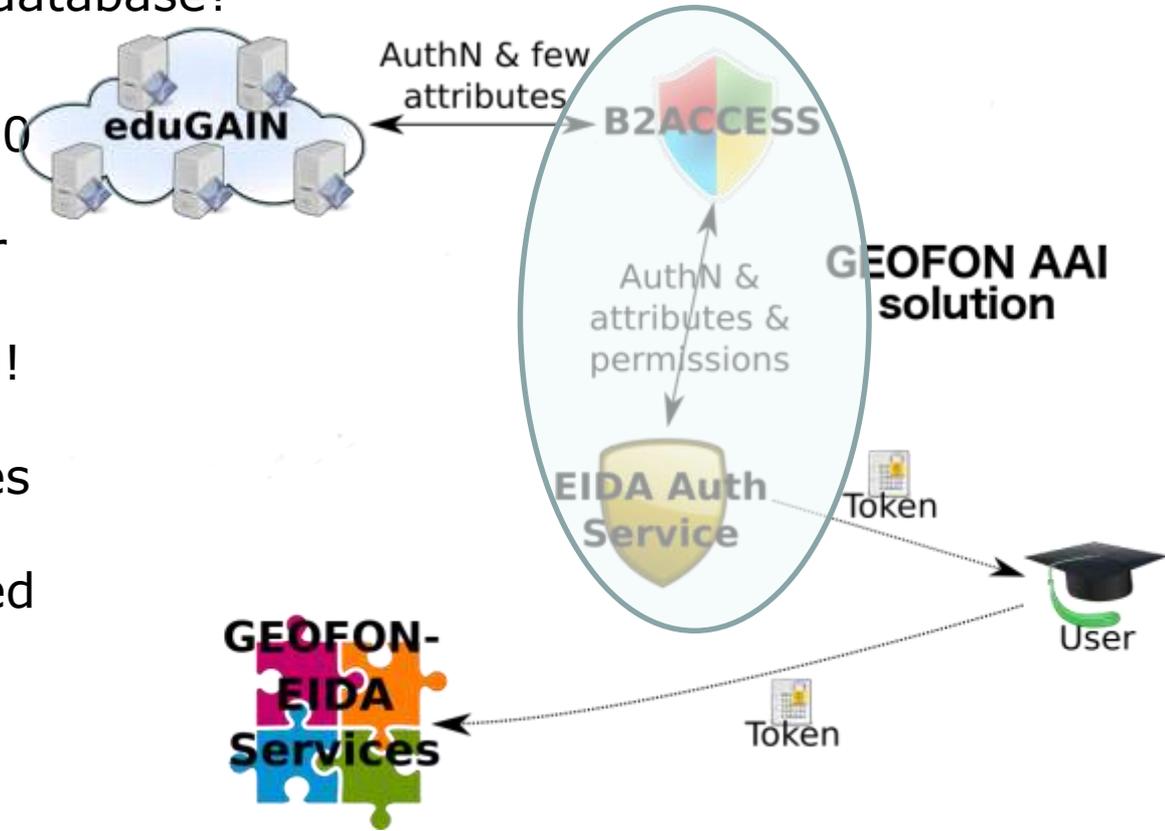
Scope: boost seismic network performance using crowdsourced triggers



Steed et al (2018). Science Advances.

Federated Authentication & Authorization

- **Challenge:** ~2000 users/year around the world, new regulations on privacy (GDPR) and services for open/embargoed data. How to properly manage a user database?
- **Approach:** Use the *eduGAIN* initiative (>2000 trustable institutions) to allow users log in at their home institutions. We don't store any user data!
- Used by many seismological data centres in Europe.
- Other groups also adopted it.
- EPOS and others also implemented this approach.



Data Management in Seismology so far...

How do you benefit from these activities (user/data provider)?

- Long term preservation.
- Data highly visible at a global scale.
- Seamless access through standard formats and protocols.
- AAI for embargoed data if needed, but also for better services in case of open data.
- Quality control on data and capability to pre-filter through metrics.
- We do the Data Management for you. You just focus on science!
- Integration into international initiatives (e.g. EPOS). We do it on your behalf.
- Data ready for Big Data projects. More users, better results.

Data Management in Seismology in the future

- Standards:
 - Data Licensing for seismic datasets (further changes for DOI).
 - Routing Service “rules” approved as an FDSN standard.
 - Metadata schema for instruments ready and presented as a DataCite standard.
- New concept for Data Management of huge volumes of seismic data (Large-N + Fibre-optic).
- Provide computation resources on top of the whole data archive (move the process towards data), but also fast Data Staging to HPC when needed.
- Provenance on the generation of event parametric data, seismic waveforms, reports, instruments metadata, with all of them interlinked.

Capacity building and outreach

Trainings – Workshops:

- ESC Training “Seismology behind the Textbooks”
 - Trieste, Italy
- Workshop/Training on Tsunami Warning Operations
 - Jakarta, Indonesia
- Deserve Winter School
 - Madaba, Jordan
- Dedicated training course for GE station operators
 - For Afgahn colleagues.
- GFZ International training course on Seismology
 - Potsdam, Germany or other country every two years (e.g. Myanmar, Ghana, Argentina).
- ORFEUS – EPOS-S Annual Meeting

Possible fields of cooperation with Armenia

- Deployment and joint operation of a new permanent multi-parametric station given a suitable location. It could become a larger multi-parametric observatory including: GPS, Seismometer, Accelerometer, Environmental sensors (wind, rainfall, temperature, humidity, pressure, etc), fluid monitoring.
- Improved data management, and processing using the in house developed software package "SeisComP3".
- Capacity building; training opportunities for Armenian scientists
- Scientific cooperation on commonly acquired data and also using other open data from the region. For interesting research topics GFZ and scientists from the region can work together to analyze the data and publish jointly the results

Possible fields of cooperation with Armenia

What can be provided by GFZ:

- Hardware and software (given a suitable site and depending on availability of funds).
- Capacity building for one of two key persons to transfer the know-how to install and operate the instruments as well as install and use software for data curation and analysis. This is fundamental to ensure sustainability.
- Support to young scientists for interesting research topics to be developed in cooperation with GFZ scientists.

What is expected from Armenia:

- Identify one or two key persons as contact points with some initial technical and scientific skills in the field of research.
- Continue the discussion to identify a suitable site (including housing) and commit to operate on the long term an eventual jointly deployed station.
- Commitment to open data policy for an eventual jointly deployed station.

Conclusions

- GEOFON (we) is continuously working towards:
 - improving the user experience;
 - facilitating access to data and its usage;
 - fostering data exchange between data centers;
 - Support for Big Data projects accessing seismological data.
- Make (you) use of our data and products as well as of the data we distribute on behalf of other partner institutions.
- If you are willing to share your data with other international organizations using standard tools and formats GEOFON can help you!
- Capacity building for local scientists in Germany or Armenia to deploy locally multi-parametric stations.
- Interested in hosting the International Training Course in Seismology? Apply to be the host on 2022!

Thank you for your attention

