The GEOFON Program in 2020

Authors: Javier Quinteros*, Angelo Strollo, Peter L. Evans, Winfried Hanka, Andres Heinloo, Susanne Hemmleb, Laura Hillmann, Karl-Heinz Jaeckel, Rainer Kind, Joachim Saul, Thomas Zieke, Frederik Tilmann.

*Phone: +49 331 288 1931. Email: javier@gfz-potsdam.de

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

Abstract

The GEOFON program consists of a global seismic network, a seismological data center and a global earthquake monitoring system. The seismic network has regional focus in Europe and North Africa as well as around the Indian Ocean, but operates stations on all continents as well as Greenland on the North American continental plate and Antarctica. The data center provides real-time seismic data through SeedLink protocol and historical data from its large archive comprising currently 120 TB of temporary and permanent seismic network data from GFZ and third party partners made available via standard services as part of the European Integrated Data Archive (EIDA) and within the International Federation of Digital Seismograph Networks (FDSN). GEOFON also provides global and rapid earthquake information. The rapid earthquake information service prioritizes fast information dissemination after moderate and large earthquakes globally, based on automatic processing. Most operations are carried out using the

SeisComP system. GEOFON distributes FAIR data, services, products and software free of charge and are used worldwide by hundreds of users and other data centers.

Introduction and historical overview

For almost three decades the GEOFON program at the German Research Centre for Geosciences (GFZ) in Potsdam (see Data and Resources) has operated both a global seismic network and a seismological data center. Potsdam can be considered the birthplace of global instrumental seismology for it was here at the Royal Prussian Geodetic Institute that in 1889 the first documented recording of a teleseismic earthquake was made by Ernst von Rebeur-Paschwitz with a pendulum horizontal seismometer, preceding a recording with an identical instrument in Wilhelmshaven by a few tens of seconds (Rebeur-Paschwitz, 1889). This initial success led Rebeur-Paschwitz as early as 1895 to propose a global network of seismometers to monitor earthquakes. The era of modern broadband seismology started in the 1970s. In Germany, the establishment of the Gräfenberg-Array (GRF) and the regional German network (GRSN), operated by the Federal Institute for Geosciences and Natural Resources (BGR) (Stammler et al., 2020), set new standards for seismic monitoring at the global level by using the newly developed high sensitivity broadband seismometers in combination with high resolution digital signal processing technology.

Shortly after its creation in 1992, GFZ established the GEOFON program to carry on the tradition of promoting global seismic monitoring and data exchange. The program evolved in synergy with the seismological community with a clear commitment towards an open data policy as well as fostering new standards in seismology right from its inception (Hanka and Kind, 1994).

Following these principles, GEOFON has fostered European cooperation in being one of the founding members of the Observatories & Research Facilities for European Seismology (ORFEUS, see Data and Resources). At the international level, the bilateral cooperation between GEOFON and the Incorporated Research Institutions for Seismology (IRIS, see Data and Resources) in 1994 was the first step to open high-quality broadband data to the seismological community, by means of the installation of a GOPHER node in Potsdam. GOPHER was the first automated global data retrieval system initiated by IRIS in the early 90s. It was based on dialup telephone lines connecting

to Data Request Manager (DRM) interfaces at the stations. Data retrieved by all GOPHER nodes (e.g. IRIS, ORFEUS, Earthquake Research Institute of Tokyo) were available to the community a few hours after an event occurred, which at that time represented a notable improvement over the status quo in the early 1990s (Hanka and Kind, 1994). Later, GEOFON played a leading role in the creation and development of the European Integrated Data Archive (EIDA), a group of data centers within ORFEUS, which were technically interconnected by software developed at GEOFON and deployed in cooperation with other European data centers, allowing users to request data from any of the data centers in a transparent way without the need to know where the data were hosted (van Eck et al., 2004; Strollo et al., this Issue).

There are major milestones to be mentioned in the history of GEOFON. First, the development and community adoption of the SeedLink protocol (see Data and Resources) in the late 1990s, which became a global de facto standard still in use today, allowed the real-time distribution of data between networks and data centers. In the early 2000s, development of the Arclink protocol (deprecated since 2019) was key to interconnecting the data archives of all EIDA data centers, and was only recently superseded by the use of FDSN web services.

A pivotal moment was the 2004 Aceh-Andaman earthquake and tsunami, which prompted the German government to fund the GITEWS project (German Indian Ocean Tsunami Early Warning System) (Lauterjung et al. 2010). In this project, GEOFON was tasked to deliver a system capable of rapidly locating and providing reliable magnitudes for potential tsunamigenic earthquakes for Indonesia. The particular challenge was that a tsunami warning system for Indonesia had to be not only effective for distant shores, the 'far-field' mode of operation for tsunami warning at the time, but also in the 'near-field' where a tsunamigenic event has to be detected within a few minutes of occurrence to notify nearby emergency service agencies in a timely manner. It quickly became clear that rapid detection, location and non-saturating magnitude determination of potentially tsunamigenic earthquakes was the only viable means to provide tsunami warnings on such a short time scale. Seismological monitoring systems existing around 2005 did not meet the requirements for a modular, distributed processing based on standard protocols. In addition, the seismic networks in operation at the time were woefully inadequate for the rapid detection of offshore earthquakes in Indonesia and most of the Indian Ocean. The GEOFON team coordinated the task of developing suitable software with the requirements of the Indonesian government regarding early warning performance targets (Hanka et al., 2010). The support by the GITEWS project led to the professionalization of software development at GEOFON and the development of SeisComP (Hanka et al., 2003; Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences and gempa GmbH, 2008) as an integrated seismic observatory software system enabling both the management of real-time seismic data flows as well as the automatic or interactive detection, location and in-depth analysis of earthquakes. In addition, GEOFON improved the seismic coverage in the Indian Ocean by establishing a backbone seismic network for the Indonesian Tsunami Early Warning System (InaTEWS, see

Data and Resources) with additional stations in other countries around the Indian Ocean, including Sri Lanka, Maldives, Yemen, Kenya and Madagascar. Besides two stations in Yemen damaged by air strikes and lost due to the ongoing war, all GEOFON's other GITEWS stations are currently in operation.

GEOFON is now positioned as one of the global reference scientific infrastructures in seismology because of the continuous extension of its global seismic network, the ingestion of growing amounts of data from partners into the data center, and the operation of a Rapid Earthquake Monitoring System (see Data and Resources), which is based on SeisComP and focuses on providing rapid location, magnitude and moment tensor information for large earthquakes worldwide. Within the next sections we present the three main GEOFON components, namely, the global seismic network, the data center, and the Rapid Earthquake Monitoring System as well as some relevant software developments.

The global seismic network

The GEOFON global seismic network presently consists of 82 high quality stations (GEOFON Data Centre, 1993). Operations started in 1993 with the installation of stations in Papua New Guinea (Port Moresby, PMG, jointly operated with IRIS and the local Geophysical Observatory), in Czech Republic (Moravsky Beroun, MORC, operated in cooperation with the local IPE-Brno and Institute of Physics of the Czech Republic) and in Ireland (Dublin, DSB, operated in cooperation with the Geophysics section of Dublin Institute for Advanced Studies), all are still in operation today through long standing cooperation with local operators and other partners. The network is among the largest global real-time seismic networks (e.g. GSN, Gee and Leith, 2011; GEOSCOPE, Institut De Physique Du Globe De Paris (IPGP) & Ecole Et Observatoire Des Sciences

De La Terre De Strasbourg (EOST), 1982), providing valuable data for almost all seismological research projects at GFZ and the wider seismological community. The GEOFON network is operated jointly with more than 50 international partners with stations on all continents, but concentrated in Europe and the Mediterranean region, as well as in the Indian Ocean (Figure 1a). Station operation is mostly performed by local partners with GFZ guidance and logistical support, allowing the global network to be technically advanced and extremely cost-effective. All stations are equipped with broadband sensors (Figure 1b); stations in regions prone to large earthquakes are additionally equipped with strong-motion sensors (GEOFON Data Centre, 1993). This setup allows resolution of the complete seismic spectrum, from small high-frequency local earthquakes to the largest global earthquakes. Recently, in collaboration with the Geodesy department of the GFZ, we augmented selected stations with GNSS and environmental sensors. One of the latest additions to the GEOFON network has been the new Accra (ACRG) multiparametric station in Ghana, equipped with broadband and strong motion sensors as well as GNSS and environmental sensors. The station was deployed during the GFZ international Training Course (ITC, see Data and resources) in October 2018 and since then successfully contributes to global monitoring filling a gap in the region. Data from all stations are acquired using various communication types (VSAT, BGAN, UMTS, DSL, etc.), and also distributed in real-time to earthquake monitoring and tsunami warning centers immediately after acquisition. During 2019 more than 100 TB of real-time data were requested from the GEOFON network (FDSN network code GE).

The data center

GEOFON operates the GFZ seismological data center, which facilitates real-time access to data from the GEOFON stations and those of many partner networks and plate boundary observatories, providing a secure long-term archive for seismological data.

GEOFON currently archives and distributes seismic data from 35 permanent networks and an increasing number of temporary networks (114 in September 2020), including those from seismological experiments carried out by scientists at German universities. It also serves as the permanent archive for the passive component of the Geophysical Instrument Pool Potsdam (GIPP) (Haberland and Ritter, 2016). A schematic view of the data center and its relation with the other GEOFON components is provided in Figure 2.

GEOFON organizes the data exchange of real-time and archived data with international partner institutions and earthquake centers. It also cooperates with them in the definition of standards of seismological data exchange within the International Federation of Digital Seismograph Networks (FDSN), ORFEUS and in the scope of bilateral initiatives (e.g. Quinteros et al., this issue). The GEOFON data center, with its current 120 TB of data holdings (Figure 3), from over 4700 globally distributed stations, but with denser coverage in the focus areas of the GFZ, is one of the largest seismological data archives in Europe.

Data management at the data center has substantially evolved in the last decade, as our focus has shifted from the initial challenge of simply acquiring high-quality seismic data in standard formats, towards better support of upstream data providers as well as provision of enhanced standard services on the downstream part, in order to enable users from the research and monitoring communities to deal with the challenges of big data.

The GEOFON data workflow distinguishes between real-time and temporary experiments. In the case of real-time data, waveforms arrive via SeedLink protocol from ~1000 stations around the world. Part of these data are used by our Rapid Earthquake Monitoring system for detection of earthquakes and to provide event parametric data within a few minutes after origin time. The full parametric data generated are stored in a database and made available to other institutions engaged in earthquake monitoring in real-time (see section HMB – http message bus) and to users via our web pages and other related services (see section The Rapid Earthquake Information). Seismic waveforms acquired in real-time from streams for which GEOFON is the authoritative data management center are ingested into our archive the day after they are acquired, while the ones used only for monitoring are kept for one year in order to be able to reproduce (or improve) our results.

In the case of temporary experiments, data providers can submit their data in miniSEED format by means of the mseed2dmc client (see Data and Resources). In parallel to the data submission, GEOFON staff work with principal investigators (PI) to generate high quality metadata. This includes the registration of the seismic network with the FDSN and the minting of a DOI (Evans et al, 2015). Once data transmission is finished, basic quality checks are performed on data and metadata. More recently the capability of streaming data from temporary experiments in real-time has been added, where generally cellular networks are used for network connectivity (e.g. Heit et al., this issue).

For all networks we calculate probabilistic power spectral density plots (PPSDs) using ObsPy (Beyreuther et al., 2010). We also support PIs to document their datasets with a technical report (e.g. Cesca et al., 2018) published by the GFZ Library and linked to the data set. To foster report publications, which is a requirement for GIPP users, we developed a convenient online tool called Report Generator (see Data and Resources), providing different templates and automated procedures to create most of the standard content required in dataset reports such as station distribution maps, station tables, and collections of PPSD plots from the available (meta)data. Recently, we included the possibility to run more detailed tests to detect problems with the (meta)data quality early by means of a semi-automatic tool called AutoStatsQ (Petersen et al., 2018), where the focus is mainly on problems with gains and orientation. This is particularly important for GEOFON, which has begun archiving of Ocean Bottom Seismometer (OBS) data from the Alfred-Wegener-Institut, Bremerhaven (AWI) and GEOMAR, Kiel. For this type of data, we follow the guidelines produced and published within the SERA project (see Data and Resources) in collaboration with many other partners.

GEOFON’s internal data policies and agreements with data providers foster the adoption of open licenses (e.g. Creative Commons CC-BY 4.0, see Data and Resources) for datasets archived at the data center. If an experiment has utilized instruments from GIPP, the PIs must agree in advance to release data under a CC-BY 4.0 license, after the end of the embargo period, which can be up to four years after the completion of the experiment (see Data and Resources). CC-BY is also the recommended license for partners using their own equipment (e.g. universities) requesting data management services from GEOFON.

The GEOFON archive is currently fully integrated into the GFZ computing center infrastructure and is ready to be scaled up in case of sudden increases in demand for storage space driven by emerging technologies such as Distributed Acoustic Sensing (DAS) applications, Large-N passive seismic experiments, etc. In addition to the two physical copies of data (mirrored, hot storage) within the GFZ campus, a third copy is replicated at the Karlsruhe Institute of Technology (KIT) using iRODS/B2SAFE services (Moore, 2008), adopted within our participation in the European projects EUDAT2020 and EOSC-hub.

All data available at the GEOFON data center are accessible through standard web services and protocols (e.g. FDSN web services, SeedLink), through a graphical interface called WebDC3 (Bianchi et al., 2015), an in-house development also running in other EIDA data centers (see Seismological Software and Data Access), and also through many widely-used client tools and libraries (e.g. ObsPy, fdsnwsscripts, Pyrocko). In Table 1 we list the services to access data and products currently provided by GEOFON. In 2019 alone, GEOFON served 200 million requests (38 TB) from the archive to more than 2700 users and 160 TB of real-time data to 300 users, including a number of operational earthquake and tsunami warning centers demanding high availability services.

Data that are still under a temporary embargo period can be accessed by means of the new federated authentication and authorization system developed at GEOFON, which has been adopted by EIDA (see Data and Resources). The system benefits from the B2ACCESS service provided by the Forschungszentrum Jülich (FZJ), which allows users from any place to be authenticated in EIDA with their existing institutional credentials (no new accounts are needed) by acting as a bridge to the eduGAIN service (see Data and Resources), which follows the same approach as the well-known eduroam Wi-Fi access authentication. This approach has many benefits considering the new European General Data Protection Regulations (see GDPR in Data and Resources), as almost no user information is kept in GEOFON systems. The PI has fine-grained control over who can access the data, and archiving at GEOFON will relieve them from having to take care of data distribution to their cooperating partners, encouraging quick archival of acquired data.

When new data are added to the archive, predefined waveform metrics such as number and length of gaps, percentage of data available, RMS, length of overlaps are calculated and stored in a database for further use in the WFCatalog system (Trani et al, 2017). This system exposes an API to query the calculated metrics, allowing the selection of data based on seismic waveform features as well as sensor geolocations and temporal specifications. Users save time by means of the system capability to provide information on the waveforms before requesting and downloading data, and data centers make better use of their connectivity resources.

All access to data and metadata are logged to produce detailed statistics. This is made available to the network operators through a Redmine content management system (see Data and Resources) used to keep updated information on the operation of all networks. The statistics are also used internally to improve our services and refine the planning of future infrastructure needs (e.g. bandwidth, storage capacity and type).

In recent years, we have gained experience not only in the establishment of sustainable data policies to support the FAIR principles (Wilkinson, et al., 2016; Findable, Accessible, Interoperable, Reusable) for the data we manage, but also in the design and implementation of workflows to improve data management.

In order to promote a culture of data citation and address the growing need to ensure that seismic networks can be reliably cited (e.g. GEOFON Data Centre, 1993), together with other data centers from the USA and Europe, we initiated and continuously foster within FDSN a set of recommendations to adopt DOIs for seismic networks and related FDSN services (see Data and Resources).

This initial effort to address and foster attribution continues within ORFEUS and other discussion groups and forums, where we are promoting usage and adoption of various persistent identifiers (PIDs) for seismological data and instrument management to enhance findability. In particular, to easily find and track all instruments in use at the GIPP we established in collaboration with the GFZ library a dedicated workflow to assign ePIC PIDs to instruments (see Data and Resources). The schema implemented in the PID record was designed as a multidisciplinary collaboration within a Working Group of the Research Data Alliance (RDA) called PIDINST, with the aim to be interoperable with other pool operators and systems from different disciplines (see Stocker et al., 2020 for details).

Our community standards ensure that seismic (meta)data are Accessible by means of open protocols (see FDSN web services in Data and Resources), and also that Datacite metadata associated with the seismic network DOI will be accessible.

All seismic (meta)data hosted at GEOFON are also Interoperable with other disciplines. A good example of this is the participation in the EPOS project, where partners from our community are working on reaching agreements to define common, standard vocabularies and ontologies following FAIR principles.

Research data are Reusable if the description of essential, recommended, and optional metadata elements are machine-readable and verifiable, its use should be easy and data should be citable to sustain data sharing and recognize the value of data (Wilkinson et al, 2016). Key concepts here are the data licensing, provenance associated with data, and usage of domain-relevant community standards. From these requirements, full provenance of data is the only point which cannot be completely fulfilled. However, an important milestone has been achieved in that we now assign well-defined licenses governing the usage of data as explained above.

In the context of further ongoing European projects, we are prototyping additional services to improve data discovery, facilitate its access, and stage data in computer facilities for later processing.

Rapid Earthquake Information

Data streams from the GEOFON network as well as ~1000 additional stations worldwide are processed in real-time to determine automatic location estimates for globally recorded earthquakes. The resulting virtual network is the so-called GEOFON Extended Virtual Network (see GEVN in Data and Resources) used to list and attribute the third-party data we use. Rapid earthquake information is provided to a number of stakeholders: earthquake and tsunami warning centers worldwide, governmental agencies, disaster management teams, news media and scientists at the GFZ and elsewhere, as well as to the general public. Earthquake parameters are distributed via e-mail, SMS, RSS news feeds, web services, web pages and GEOFON forum. A popular web page is the Global Seismic Monitor (see Data and Resources) showing the latest largest events in the world. On average ~5000 events and ~1000 moment tensor solutions are published per year (Figure 4) and disseminated to the various stakeholders according to their user profiles. GEOFON is a key nodal member of the European-Mediterranean Seismological Centre (EMSC, Godey et al., 2006). All processing is based on the SeisComP software package (Helmholtz-Centre Potsdam-GFZ German Research Centre for Geosciences & gempa GmbH, 2008). The GEOFON rapid earthquake information is among the first disseminated in the aftermath of large earthquakes worldwide. Events are published with minimal delay to the web page, GEOFON forum and RSS feeds whereas sms and email alerts dissemination takes place within typically ~10 minutes (Figure 5).

Aiming at reducing the publication delay we teamed up with EMSC and the Hungarian Academy of Sciences and developed a novel approach for early detection of felt events (Steed et al, 2019) that combines crowdsourced data from EMSC with our automatic seismic detections. In the initial test data set this approach reduced the publication delay from 3 minutes to 1.7 minutes (median value). This development has been facilitated by the development of the HMB software (Heinloo, 2016) that allowed GEOFON to stream real-time parametric information to EMSC. In addition, GEOFON also provides high quality moment tensor solutions for major earthquakes (Figure 4c and 4d).

Seismological Software

The GEOFON program has a long history of open software development for the community. In many cases, software products developed here have been broadly adopted by the users. Nowadays, GEOFON provides not only high quality software for seismological purposes, but also for topics like high performance communication systems and data management. Here we mention only a selection of our software products.

SeisComP

The GEOFON operations, spanning from the remote seismic stations to the rapid earthquake information through the data center are based on the SeisComP® software package (Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences and gempa GmbH 2008). SeisComP features waveform data acquisition, distribution, quality control, archival and analysis in real-time. Graphical user interfaces are available for configuration, manual event revision, event display and state-of-health monitoring (Figure 6a and b).

The software has been developed within the GFZ since 2005 as part of the GITEWS project, as described in the Introduction, and since 2009 a newly founded spin-off company of the GFZ, gempa GmbH, started to further develop the package, together with tailor made solutions for users. Several important additional modules were either developed by external contributors or funded by them. Because of its accessible GUI, powerful back-end and the fact that its use was free for non-commercial users it has become a widely-used package for seismological data acquisition, data processing and data exchange. Within the context of several major Pan-European infrastructure projects coordinated by ORFEUS (e.g. MEREDIAN, NERIES, NERA), major developments in European seismology, in particular the VEBSN (Virtual European Broadband Seismological Network), EIDA (European Integrated Data Archive) and RRSM (Rapid Strong Motion Portal) could be achieved more easily by building on the SeisComP software platform, which is in use at most European data centers. In the context of the GITEWS project, the software architecture was from the onset designed to meet the stringent requirements of a 24/7 tsunami early warning center. Since the first public release in 2008 more than 500 SeisComP Public Licenses (SPL) have been released to institutions world-wide. This includes many earthquake and tsunami warning operational centers e.g. ETHZ-SED (Switzerland), CEA (France), GNS (New Zealand), GA (Australia), and BMKG (Indonesia), which also contributed to further develop SeisComP in collaboration with gempa.

In April 2020, SeisComP (version 4) was released with major upgrades in core software components and more comprehensive documentation. The most important change has been the move from the SPL license to the standard GNU Affero Public License (AGPL), which now allows both non-commercial and commercial users to freely use the software without the need to apply for a license, and makes it more attractive for seismological software developers to contribute to the code-base or develop plug-ins.

WebDC3

WebDC3 (Bianchi et al., 2015) is a web interface working on top of the FDSN standard web services (Figure 6c). It is AJAX-based and allows users to conveniently discover seismic stations and explore events in seismic catalogues, build and submit requests for data and metadata and finally download the results in different formats. Requests can be built using either absolute time windows or by station-event combinations suitable for different data processing pipelines. Furthermore, construction of data requests and the actual download can be separated, allowing for larger requests, in particular for users with low bandwidth or unstable connections.
Two operational modes are available, either as a front end to a single data center offering their data holdings, or as a front end to a federation of data centers, giving access to data holdings from all data centers federated by a Routing Service instance (see Routing Service below).
WebDC3 is available as an open-source package and has been adopted by a number of seismic agencies including KNMI (Netherlands), SED (Switzerland), NIEP (Romania), NOA (Greece), KOERI (Turkey), USP (Brazil), and BMKG (Indonesia).

Fdsnws scripts

fdsnwsscripts (see Data and Resources) is a collection of distributed data request clients that are based on FDSN web services (fdsnws-station, fdsnws-dataselect) and the EIDA Routing Service

(Quinteros, 2017). The client tools in this development provide seismic waveform data in miniSEED format and seismic metadata in Station XML.

The main tool of the package is fdsnws_fetch, conceived as a straightforward replacement for the classic arclink_fetch, which was a distributed request tool making use of the Arclink protocol. This tool not only provides backward compatibility, but also supports all new services and functionalities offered within the EIDA and FDSN ecosystems. For user convenience and easier integration into existing workflows, it accepts various legacy input formats, such as arclink_fetch and BREQ_FAST. Based on the request content, in addition to delivering data and metadata, the formal citation string including authors and DOIs to be included in manuscripts is provided to the user, thus encouraging wider adoption of recommended data citation practices.

Other types of request clients are also included in the package. For instance, a specialized client, fdsnws2sds, supporting requests for very large continuous datasets (including the capability to resume downloads) and another one, fdsnws2seed, providing data in the legacy SEED format.

Routing Service

The Routing Service (Quinteros, 2017) is a service to federate data centers by integrating multiple sources offering access to data and products using compatible types of services. This service is a core service of the current EIDA infrastructure, a schematic view of the service is provided in Figure 6d.

Examples of standard data and product objects are seismic time-series waveforms and station inventories, but operators can optionally include other services with a similar interface (stream codes and time windows), for instance WFCatalog, the web service providing quality control metrics, which has been adopted by all EIDA data centers (Trani et al, 2017).

This routing service is supported by many popular clients in the community, like ObsPy (Beyreuther et al., 2010), fdsnws_fetch and WebDC3 among others.

HMB – http message bus

HMB (Heinloo, 2016) is a general-purpose message bus, which has been designed to be compatible with SeisComP. It has been developed to distribute objects, such as event detections, in real-time. HMB supports any objects that can be expressed in JSON or BSON formats, including all SeisComP data model objects and even binary data, such as waveform records. HMB is based on HTTP and works with all standard web servers. It is the core of the real-time exchange process between GEOFON and other institutions (e.g. EMSC, BMKG, USGS-NEIC) regarding parametric information (locations, magnitudes and amplitudes) from our monitoring system.

Conclusions and future challenges

Since its inception, GEOFON has been promoting the open distribution of seismological data and software as well as the development and adoption of standards, following the FDSN guiding principles. Decades of efforts in this direction paved the road to data FAIRness in seismology at GEOFON.

The GEOFON Global Seismic network, initially used as a vehicle for networking in order to improve open data exchange and standardization, is today our backbone network for real-time monitoring and for technical developments. Extension towards other observables (GNSS, environmental parameters) is ongoing at selected stations, and their integration, in addition to the modernization of legacy instruments, will remain a challenging task for the decade to come.

The GFZ seismological data center operated by GEOFON offers a secure place for GFZ's own data and those of its partners, with data policies that implement the recommendations of international initiatives within and beyond seismology (e.g., FDSN, RDA). The exponential growth of data volume driven by technological innovation is a challenge that we will need to address in the near future. Leveraging on EU funded projects, we are developing novel data management strategies for emerging technologies such as DAS, cost effective micro sensors for the Internet of Things (IoT), and other devices enabling dense observations. This is not only challenging in the context of existing seismological data curation and distribution paradigms (Quinteros et al., this issue), but also needs to be taken into account during data collection where new efficient acquisition and transmission protocols are also needed. The SeedLink transmission protocol will need to be enhanced to cope with these new needs, and discussions on this have already begun.

Rapid earthquake information distributed by GEOFON are among the fastest available after large earthquakes worldwide. The parametric earthquake information is distributed in real-time to various stakeholders spanning from tsunami warning centers to the general public. We are working towards extending the offer of earthquake related products, leveraging on internal expertise available at GFZ, both technical and scientific.

The competences gained through the years in seismological software development for both real-time applications and data management have enabled us to make progress with the challenges listed above. In addition, the excellent networking within and outside GFZ and Germany will facilitate integration and cooperation with other institutions and topics.

We remain engaged in knowledge and technology transfer as we consider this an important component of our service portfolio to ensure sustainability of our activities. We do this in synergy with the GFZ International Training Course and participate in a number of additional outreach and/or training events. We also work closely together with the Public Relations department of the GFZ to improve public understanding of earthquakes, in particular after damaging or noteworthy events, and support the school lab of the GFZ to produce learning materials (e.g., a recent outreach product showing global seismicity, focal mechanisms and hazard).

The GEOFON Program, initially inspired by the idea of Rebeur-Paschwitz proposing a global network of seismometers to monitor earthquakes, has drastically evolved through the years since inception. The program began after the first disruptive wave of changes in seismology of the 1980’s and played a lead role in important developments for the global and European communities. The program’s work on SeedLink, SeisComP, and EIDA provided the technical know-how to realize the first prototype of the modern seismological “Cloud”. Seismology and geoscience in general are currently going through the second disruptive wave of changes with Big Data and Artificial Intelligence. FAIR GEOFON data and services facilitate the work of the next generation of scientists to unveil new scientific discoveries that might be hidden in the massive data holdings spanning several decades.

Data and Resources

All links in this section were last accessed in October 2020.

- The GEOFON Program home page: https://geofon.gfz-potsdam.de/
- ORFEUS home page can be found in https://www.orfeus-eu.org/.
- IRIS home page can be found in https://www.iris.edu/.
- InaTEWS: https://inatews.bmkg.go.id/en_index.php

- The GEOFON Rapid earthquake monitoring system is available at http://geofon.gfz-potsdam.de/eqinfo/.
- Documentation about the SeedLink protocol and server can be found at https://www.seiscomp.de/doc/apps/seedlink.html.
- International Training Course: https://www.gfz-potsdam.de/en/about-us/education-and-training/seismology-and-hazard-assessment/.
- miniseed2dmc client can be found at https://github.com/iris-edu/miniseed2dmc
- Detailed information about the different Creative Commons Licenses can be found in https://creativecommons.org/licenses/.
- The Report Generator can be found in its Github repository at https://github.com/rizac/gfzreport.
- The SERA Deliverable "Metadata challenges and proposed solutions" by Clinton, et al. (2018) can be downloaded from http://www.sera-eu.org/export/sites/sera/home/.galleries/Deliverables/SERA_D4.2_Metadata-challenges-and-proposed-solutions.pdf.
- GIPP terms of use: https://www.gfz-potsdam.de/fileadmin/gfz/sec22/pdf_doc/GIPP/rules_GIPP_10_2019.pdf
- The Federated EIDA Authentication-Authorization software can be accessed at https://geofon.gfz-potsdam.de/eas.
- Information about the B2ACCESS service can be found at https://www.eudat.eu/services/b2access.
- Information about eduGAIN, the interfederation service that connects identity federations around the world, can be found at https://edugain.org/.

- Information on the current General Data Protection Regulation can be found at https://gdpr.eu/.
- Redmine software and documentation can be seen at https://www.redmine.org/.
- The “FDSN recommendations for seismic network DOIs and related FDSN services” has been published with doi: 10.7914/D11596.
- Information about the Handle system can be found in https://www.handle.net/hnr_documentation.html.
- Information about ePIC PIDs can be found in https://www.pidconsortium.net/.
- Specifications of the FDSN web services can be found at http://www.fdsn.org/webservices/FDSN-WS-Specifications-1.2.pdf.
- Detailed information about the GEOFON Extended Virtual Network (GEVN) can be read at https://geofon.gfz-potsdam.de/eqinfo/gevn/.
- Information about the B2SAFE service can be found at https://www.eudat.eu/services/b2safe.
- Information of the EOSC-hub project (Integrating and managing services for the European Open Science Cloud) and for the EUDAT project (Collaborative Data Infrastructure) are available at https://www.eosc-hub.eu/ and https://eudat.eu/ respectively.
- Documentation about the Rapid Raw Strong Motion Portal (RRSM) is available at https://www.orfeus-eu.org/opencms/rrsm/information/.
- Information about fdsnwsscripts can be found in https://geofon.gfz-potsdam.de/software/fdsnws_scripts/.
- Detailed and updated documentation of the Routing Service is publicly available at https://routing.readthedocs.io/.

- Poster for the School lab (in German) can be found at https://www.gfz-potsdam.de/en/media-and-communication/current-earthquake-information/.

Acknowledgements

GEOFON's success in becoming a key source of data and in providing accurate reliable locations would be impossible without the many partners supporting the installation and ongoing operation of the global seismic network, contributing data to the archive, and making a dense global network of stations available for real-time processing. We thank in particular all partner’s co-operating stations, the network operators making their data available in real-time and thus contributing to the GEVN, and all data providers who decided to let GEOFON curate and distribute their data. We thank also former GEOFON colleagues who contributed in one way or the other to the Program: Bernd Weber, Jan Becker, Marcelo De Bianchi, Juergen Schuette, Wolfgang Kohl, Martin Haxter, Michael Guenther, Mathias Hoffmann, Oliver Backmann, Riccardo Zaccarelli. Although the GEOFON program operation is funded with core budget, numerous third party projects have supported many of the developments of the GEOFON data centre and network. For the last five years, we thank in particular the EU projects (next to the project acronyms the Grant Agreement number): EOSC-Pillar (857650), RISE (821115), EOSC-hub (777536), SERA (730900), EPOS-IP (676564), EUDAT2020 (654065), as well as the BMBF projects EWRICA and GeoDataNode. The initial development of SeisComP has been triggered by the GITEWS project funded by BMBF and implemented at BMKG. We acknowledge the Library and Information Services and Computing Centre of the GFZ for the support provided in data management as well as the Karlsruhe Institute of Technology (KIT) and Forschungszentrum Jülich (FZJ) for providing off site data replication capacity and authentication services respectively. GEOFON is operated by the Seismology section at the GFZ and is part of GFZ's Modular Earth Science Infrastructure (MESI). Finally, we thank F. Haslinger, K. Stammler, W. Friederich, T. Meier, M. Wyss, J. Zednik and former members for serving on the advisory committee.

References

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python Toolbox for Seismology, Seismol. Res. Lett., doi: 10.1785/gssrl.81.3.530.

Bianchi, M., P. L. Evans, A. Heinloo, and J. Quinteros (2015). WebDC3 Web Interface, GFZ Data Services, doi: 10.5880/GFZ.2.4/2016.001.

Bird, P. (2003). An updated digital model of plate boundaries, G-cubed, AGU, 4. doi: 10.1029/2001gc000252.

Cesca, S., M. Sobiesiak, A. Tassara, M. Olcay, E. Günther, S. Mikulla, and T. Dahm (2018). The Iquique Local Network and PicArray, GFZ Data Services, doi: 10.2312/gfz.b103-18022.

Evans, P. L., A. Strollo, A. Clark, T. Ahern, R. Newman, J. F. Clinton, H. Pedersen, and C. Pequegnat (2015). Why seismic networks need digital object identifiers, Eos, 96, doi: 10.1029/2015EO036971.

Gee, Lind S. and W. S. Leith (2011). The Global Seismographic Network. doi: 10.3133/fs20113021.

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

Godey, S., R. Bossu, J. Guilbert, and G. Mazet-Roux (2006). The Euro-Mediterranean Bulletin: A Comprehensive Seismological Bulletin at Regional Scale, Seismol. Res. Lett. 77, 460-474.

Haberland, C. and O. Ritter (2016). GIPP: Geophysical Instrument Pool Potsdam, Journal of large-scale research facilities, 2, A64, doi: 10.17815/jlsrf-2-128.

Hanka, W. and R. Kind (1994). The GEOFON Program, Annals of Geophysics, 37, 1060-1065. doi: 10.4401/ag-4196.

Hanka, W., A. Heinloo, and K.-H. Jaeckel (2003). Networked Seismographs: GEOFON RealTime Data Distribution, In: Proceedings of OHP/ION symposium “LongTerm Observations in the Oceans: Current Status and Perspectives for the Future”, January 21–27, 2001, Workshop Report, 58–62.

Hanka, W., J. Saul, B. Weber, J. Becker, P. Harjadi, Fauzi, and GITEWS Seismology Group (2010). Real-time earthquake monitoring for tsunami warning in the Indian Ocean and beyond, Nat. Hazards Earth Syst. Sci., 10, 2611-2622, doi: 10.5194/nhess-10-2611-2010, 2010.

Heinloo, A. (2016). The httpmsgbus general-purpose HTTP-based message bus, GFZ Data Services, doi: 10.5880/GFZ.2.4.2016.001.

Heit, B., L. Cristiano, C. Haberland, F. Tilmann, D. Pesaresi, Y. Jia, H. Hausmann, S. Hemmleb, M. Haxter, T. Zieke, K.-H. Jaeckl, A. Schloemer, and M. Weber (2020). The SWATH-D seismological network in the Eastern Alps, Seismol. Res. Lett., submitted to this issue.

Helmholtz-Centre Potsdam-GFZ German Research Centre for Geosciences & gempa GmbH. (2008). The SeisComP seismological software package. GFZ Data Services. doi: 10.5880/GFZ.2.4.2020.003.

Institut De Physique Du Globe De Paris (IPGP) and Ecole Et Observatoire Des Sciences De La Terre De Strasbourg (EOST) (1982). GEOSCOPE, French Global Network of broad band seismic stations. Institut de physique du globe de Paris (IPGP), Université de Paris. doi: 10.18715/GEOSCOPE.G

Lauterjung, J., P. Koltermann, U. Wolf, and J. Sopaheluwakan (2010). The UNESCO-IOC framework - establishing an international early warning infrastructure in the Indian Ocean region, Nat. Hazards Earth Syst. Sci., 10, 2623–2629, doi: 10.5194/nhess-10-2623-2010.

Moore, R. W. (2008). Towards a Theory of Digital Preservation, Int. Jour. Digital Curation, Issue 1, vol 3, 63-75.

Petersen, G. M., S. Cesca, M. Kriegerowski, and the AlpArray Working Group (2018). Automated Quality Control for Large Seismic Networks: Implementation and Application to the AlpArray Seismic Network, Seismol. Res. Lett., 90, 3, 1177-1190, doi: 10.1785/0220180342.

Quinteros, J. (2017). Routing Service: A data centre federation for the seismological community, GFZ German Research Centre for Geosciences, doi: 10.5880/GFZ.2.4/2017.001.

Quinteros, J., J. A. Carter, J. Schaeffer, C. Trabant, and H. Pedersen (2020). Changing Use Paradigms for Seismological Data Repositories, Seismol. Res. Lett., submitted to this issue.

Rebeur-Paschwitz, E. (1889). The Earthquake of Tokio, April 18, 1889, Nature, 40, 1030, p 294-295.

Stammler, K., M. Bischoff, A. Brüstle, L. Ceranna, S. Donner, K. Fischer, P. Gaebler, W. Friederich, S. Funke, G. Hartmann, B. Homuth, B. Knapmeyer-Endrun, M. Korn, T. Megies, C. Pilger, T. Plenefisch, I. Pustal, I. Rappsilber, B. Schmidt, L. Sonnabend, S. Stange, J. Wassermann, U. Wegler (2020). German Seismic and Infrasound Networks contributing to the European Integrated Data Archive (EIDA), Seismol. Res. Lett., submitted to this issue.

Steed, R. J., A. Fuenzalida, R. Bossu, I. Bondar, A. Heinloo, A. Dupont, J. Saul, and A. Strollo (2019). Crowdsourcing triggers rapid, reliable earthquake location, Science Advances, 5, 4, eaau9824, doi: 10.1126/sciadv.aau9824.

Stocker, M., L. Darroch, R. Krahl, T. Habermann, A. Devaraju, U. Schwardmann, C. D'Onofrio, and I. Häggström (2020). Persistent Identification of Instruments, Data Sci. J., 19(1), p.18. doi: 10.5334/dsj-2020-018.

Strollo, A. , D. Cambaz, J. Clinton, P. Danecek, C. Evangelidis, A. Marmureanu, L. Ottemoller, H. Pedersen, R. Sleeman, K. Stammler, D. Armbruster, J. Bienkowski, K. Boukouras, P. Evans, M. Fares, C. Neagoe, S. Heimers, A. Heinloo, M. Hoffmann, P. Kaestli, V. Lauciani, J. Michalek, E. O. Muhire, M. Ozer, L. Palangeanu, C. Pardo, J. Quinteros, M. Quintiliani, J. A. Jara Salvador, J. Schaeffer, A. Schloemer and N. Triantafyllis (2020). EIDA: the European Integrated Data Archive and service infrastructure within ORFEUS, Seismol. Res. Lett., submitted to this issue.

Trani, L., M. Koymans, M. Atkinson, R. Sleeman, and R. Filgueira (2017). WFCatalog: A catalogue for seismological waveform data, Comput. Geosci., 106, 101-108, doi: 10.1016/j.cageo.2017.06.008.

van Eck, T., C. Trabant, B. Dost, W. Hanka, and D. Giardini (2004). Setting up a virtual Broadband Seismograph Network across Europe, Eos Trans. AGU, 85, 13.

Wilkinson, M., M. Dumontier, I. Aalbersberg, et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship, Sci. Data 3, 160018, doi: 10.1038/sdata.2016.18.

Full mailing address for authors

Javier Quinteros: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. javier@gfz-potsdam.de ; orcid: 0000-0001-9993-4003

Angelo Strollo: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. strollo@gfz-potsdam.de orcid: 0000-0001-9602-6077

Peter L. Evans: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. pevans@gfz-potsdam.de

Winfried Hanka: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. hanka@gfz-potsdam.de

Andres Heinloo: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. andres@gfz-potsdam.de

Susanne Hemmleb: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. hemmleb@gfz-potsdam.de ; orcid: 0000-0003-1516-0392

Laura Hillmann: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. laura.hillmann@gfz-potsdam.de ; orcid: 0000-0002-2648-2365

Karl-Heinz Jaeckel: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. karl-heinz.jaeckel@gfz-potsdam.de

Rainer Kind: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. kind@gfz-potsdam.de

Joachim Saul: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. saul@gfz-potsdam.de

Thomas Zieke: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. thomas.zieke@gfz-potsdam.de

Frederik Tilmann: GFZ German Research Centre for Geosciences, Albert-Einstein-Straße 42-46, 14473 Potsdam, Germany. tillmann@gfz-potsdam.de orcid: 0000-0002-7439-8782

Tables

Table 1: List of GEOFON services with name, entry point and description.

Name	Location / entry point	Description
fdsnws-station	http://geofon.gfz-potsdam.de/fdsnws/station/1/	Provide metadata information for the data in archive in text and XML formats (EIDA service)
fdsnws-dataselect	http://geofon.gfz-potsdam.de/fdsnws/dataselect/1/	Waveform data provisioning system (archive) in mseed format (EIDA service)
SeedLink	geofon.gfz-potsdam.de:18000	Real-time waveform data provisioning system in mseed format
WFCatalog	http://geofon.gfz-potsdam.de/eidaws/wfcatalog/1/	Quality metrics of seismic waveforms in archive (EIDA service)
WebDC3	http://eida.gfz-potsdam.de/	Graphical User Interface (GUI) to access GFZ and EIDA data holdings by

		events or absolut time
AAI	https://geofon.gfz-potsdam.de/eas	Authentication service (EIDA service)
EQinfo	https://geofon.gfz-potsdam.de/eqinfo/	Interactive access to Rapid Earthquake Information
fdsnws-event	http://geofon.gfz-potsdam.de/fdsnws/event/1/	Event web service to access Rapid Earthquake Information (parametric data contributed to EMSC and ISC)
Routing Service	http://geofon.gfz-potsdam.de/eidaws/routing/1/	Routes to GFZ data holdings and synchronization (EIDA service)
GEOFON forum	https://geofon.gfz-potsdam.de/forum/	Forum with News, updates on Big and XXL events and other topics
Redmine	https://geofon.gfz-potsdam.de/redmine/	GEOFON Program management and issue tracking tool (for stations

		and network operators only)

Figure captions

Figure 1: a) GEOFON real-time seismic stations. Red triangles indicate currently operational stations, while former non-operational stations are indicated with reversed triangles in grey; b) example of a GEOFON station (GE.SALP, Salfit, Palestinian territories). Left: solar power system with the VSAT antenna used as primary communication channel. Right: acquisition rack with digitizer, rugged PC to buffer data locally, VSAT modem, mobile router, controller for power management and SOH. In the center the broadband sensor.

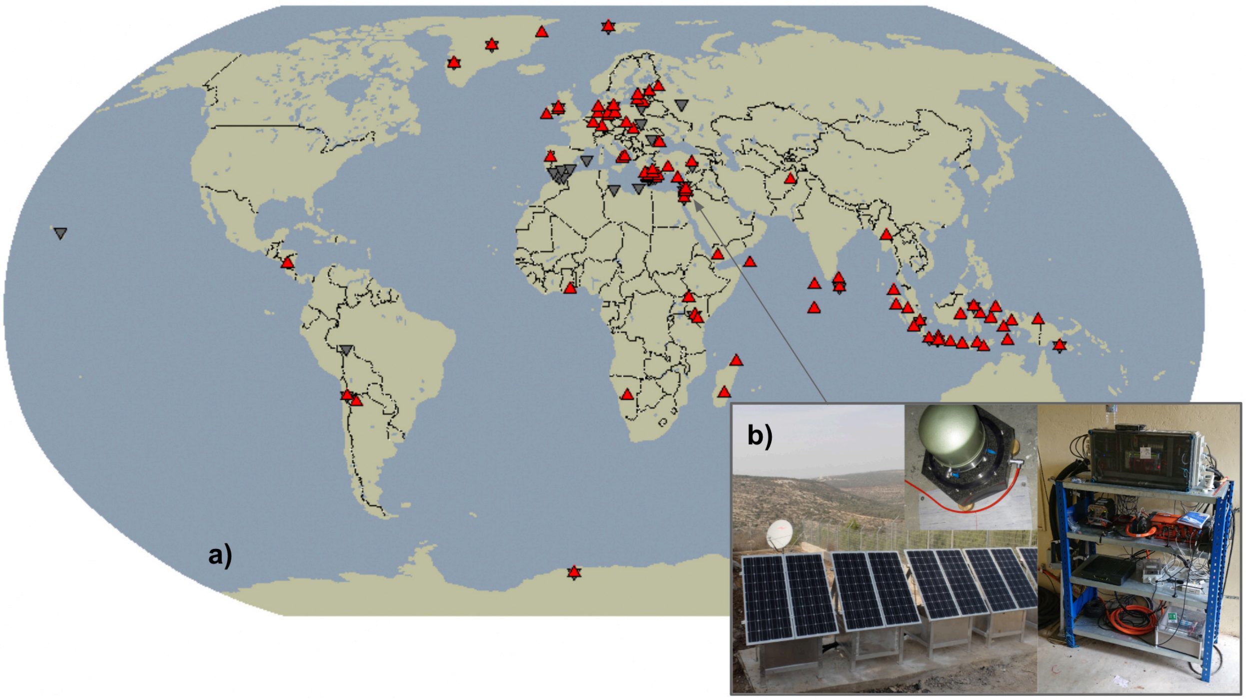
Figure 2: Schematic view of the GEOFON data center and its relation with the other GEOFON components. Data from the GE network are flowing to the archive where they are curated and used with all other real-time streams of the GEVN for the rapid earthquake information. Real-time and historical data are served to users by means of SeedLink, FDSN web services and WebDC3 between others.

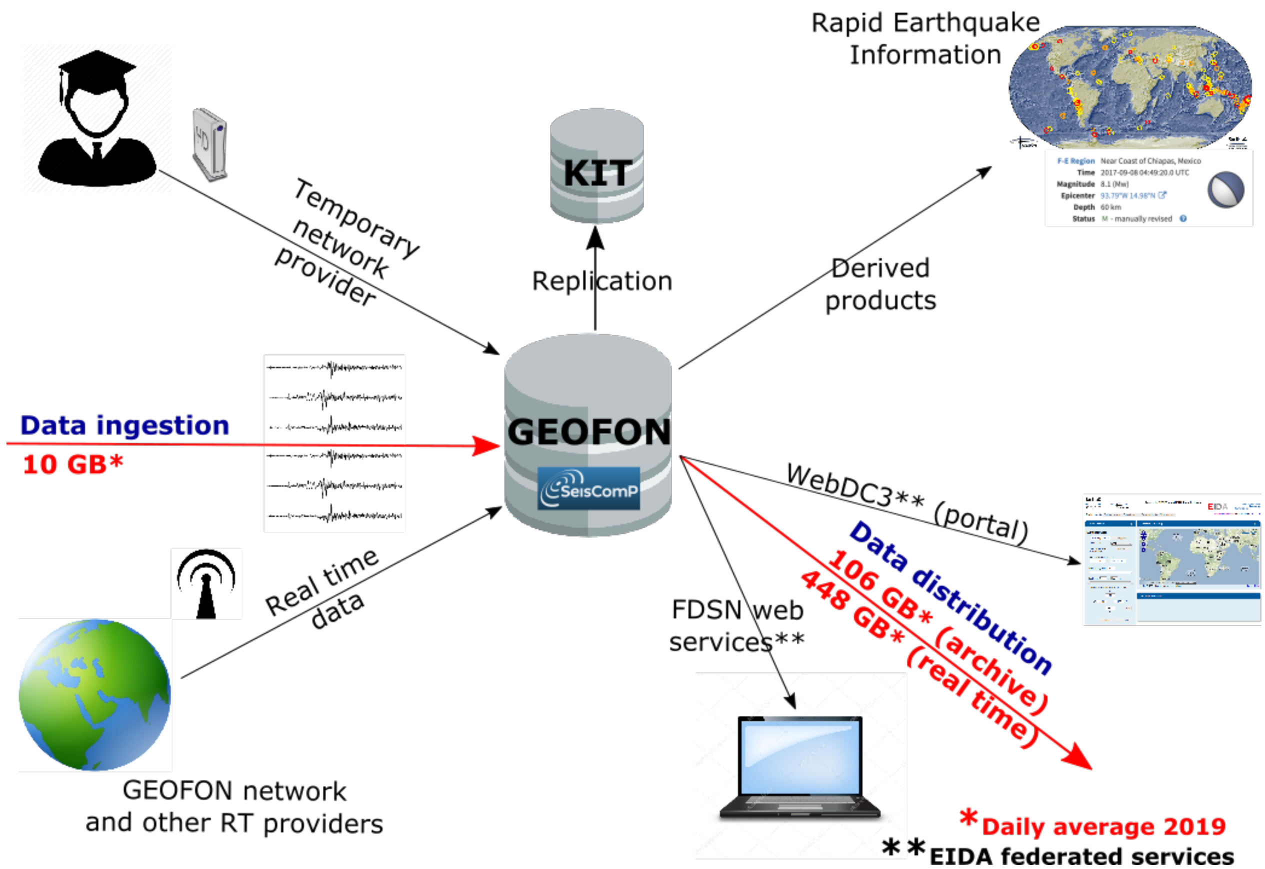
Figure 3: a) distribution of the approximately 4790 stations archived at GEOFON (October 2020). Permanent stations are plotted with inverted triangles, while temporary stations are shown with normal triangles. b) data archive size by year of data acquisition and type of experiment (permanent or temporary networks). The curve plotted in red indicates the fraction of strong-motion data.

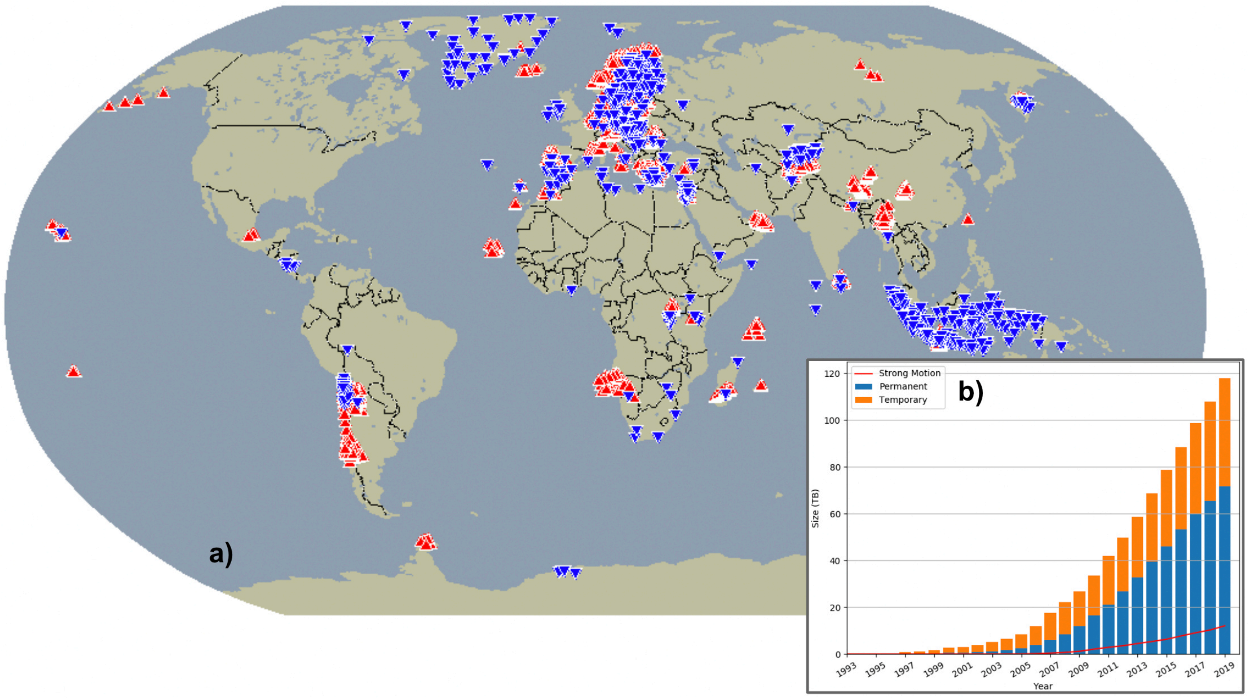
Figure 4: a) Distribution of earthquake locations published by GEOFON (2008-2019) by magnitude (size) and depth (color). b) Frequency magnitude distribution for all events published 2016-2019. The line shows the cumulative distribution, the histogram the number of earthquakes in each 0.1 magnitude bin. Revised means that a human analyst has either confirmed this event or added or removed picks manually. c) Moment tensors calculated by GEOFON 2011-2019 (only Mw>=5.5 events shown for clarity). Red: thrust events, Green: Strike-slip, Blue: normal, Gray: oblique. Yellow-orange line shows plate boundaries after Bird (2003). d) Frequency magnitude distribution for all moment tensors published 2016-2019. The line shows the cumulative distribution, the histogram the number of earthquakes in each 0.1 magnitude bin.

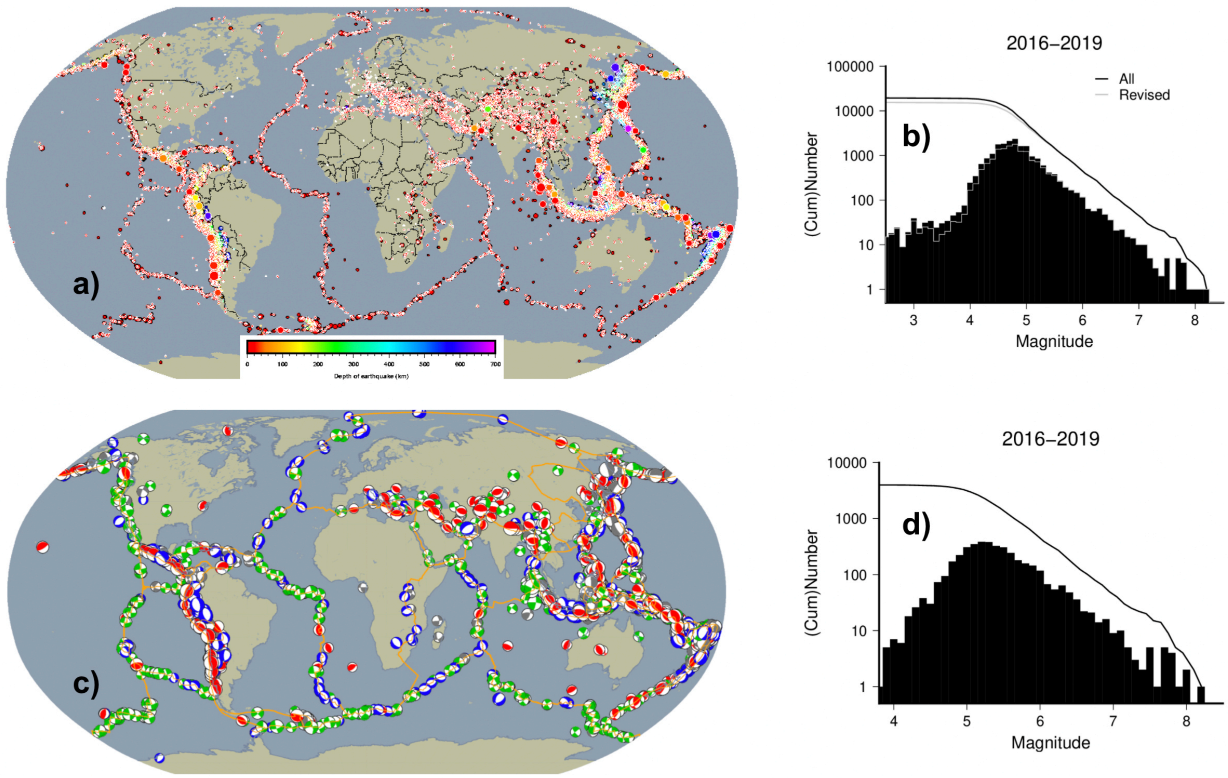
Figure 5: Time elapsed since origin time for first automatic publication of events on the web site (blue) and for alerting subscribed users (green) via SMS, Email or the GEOFON Forum. Numbers are for 2019.

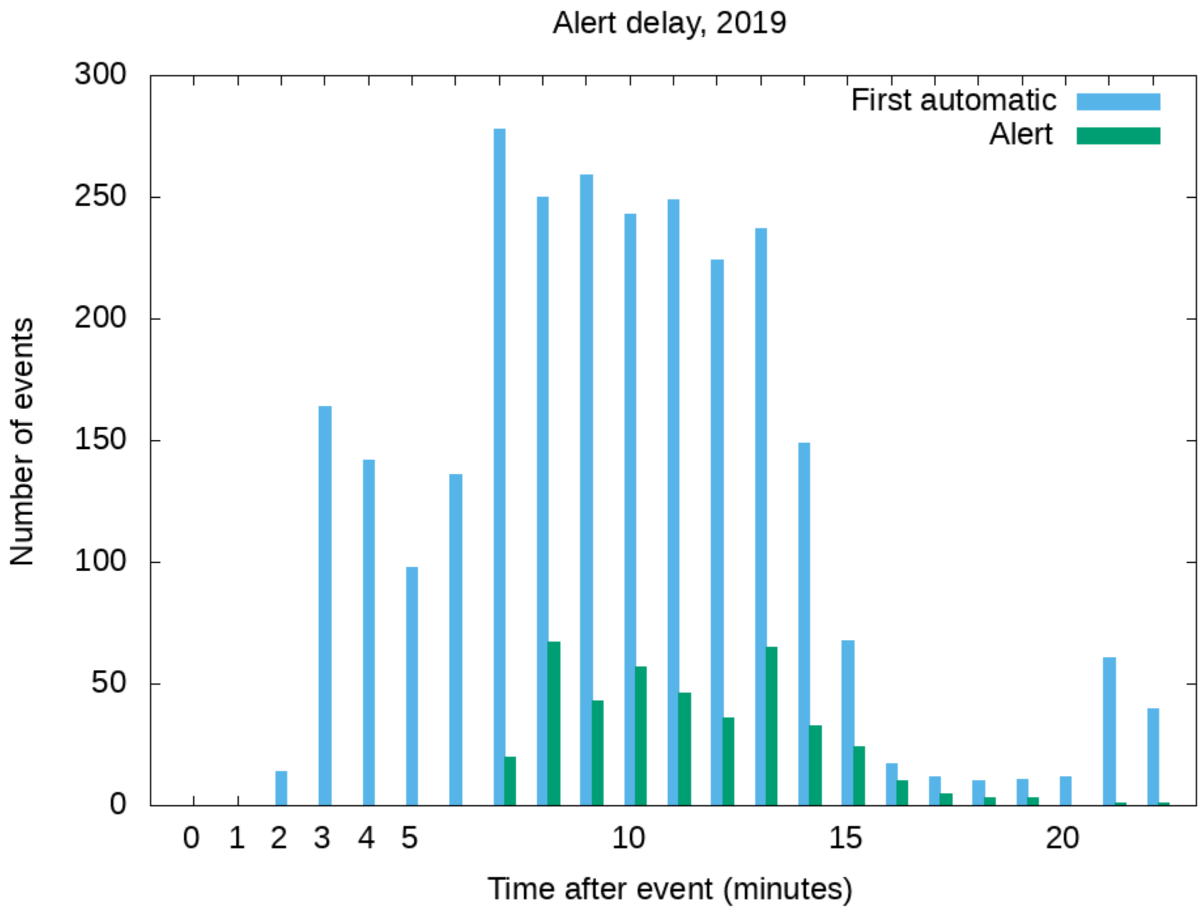
Figure 6. Screenshots of modules of the SeisComP software: a) origin locator view (scolv); b the picker. c) Screenshot of the WebDC3 software and d) schema of the routing service.

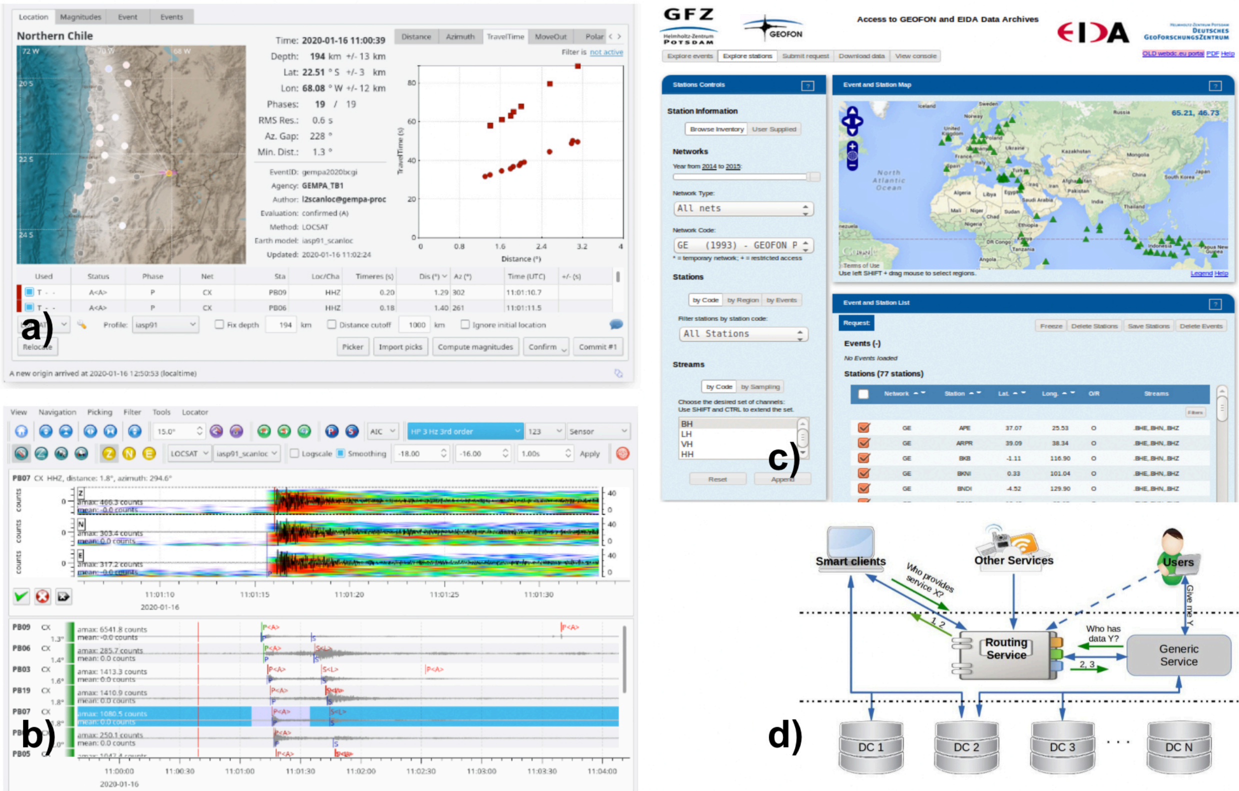

a)
b)


Rapid Earthquake Information
KIT
Temporary network provider
Replication
Derived products
Data ingestion
10 GB*
GEOFON
SeisComP
WebDC3** (portal)
Data distribution
106 GB* (archive)
448 GB* (real time)
Real time data
FDSN web services**
GEOFON network and other RT providers
*Daily average 2019
**EIDA federated services


a)
b)
Strong Motion
Permanent
Temporary
Size (TB)
Year
0
20
40
60
80
100
120
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011
2013
2015
2017
2019


a)
Depth of earthquake (km)
0
100
200
300
400
500
600
700
b)
2016–2019
All
Revised
(Cum)Number
Magnitude
c)
d)
2016–2019
(Cum)Number
Magnitude


Alert delay, 2019
First automatic
Alert
Number of events
300
250
200
150
100
50
0
0 1 2 3 4 5 10 15 20
Time after event (minutes)


a)
Northern Chile
b)
c)
Access to GEOFON and EIDA Data Archives
d)
Smart clients
Other Services
Users
Who provides service X?
Routing Service
Who has data Y?
Generic Service
DC 1
DC 2
DC 3
DC N