EIDA: the European Integrated Data Archive and service infrastructure within ORFEUS

Authors

Angelo Strollo[1], Didem Cambaz[2], John Clinton[3], Peter Danecek[4], Christos P. Evangelidis[5], Alexandru Marmureanu[6], Lars Ottemöller[7], Helle Pedersen[8], Reinoud Sleeman[9], Klaus Stammler[10], Daniel Armbruster[3], Jarek Bienkowski[9], Kostas Boukouras[5], Peter L. Evans[1], Massimo Fares[4], Cristian Neagoe[6], Stefan Heimers[3], Andres Heinloo[1], Matthias Hoffmann[10], Philippe Kaestli[3], Valentino Lauciani[4], Jan Michalek[7], Erich Odon Muhire[10], Mehmet Ozer[2], Lucian Palangeanu[6], Constanza Pardo[11], Javier Quinteros[1], Matteo Quintiliani[4], Jose Antonio Jara-Salvador[12], Jonathan Schaeffer[13], Antje Schloemer[14], Nikolaos Triantafyllis[5]

Affiliation

[1] Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany

[2] Bogazici University, Kandilli Observatory and Earthquake Research Institute, Istanbul, Turkey

[3] Swiss Seismological Service, ETH Zurich, Zurich, Switzerland.

[4] Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

[5] National Observatory of Athens, Institute of Geodynamics, Athens, Greece

[6] National Institute for Earth Physics (NIEP), Romania

[7] Department of Earth Science, University of Bergen, Norway

[8] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France

[9] Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

[10] Federal Institute for Geosciences and Natural Resources (BGR), Germany

[11] Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

[12] Institut Cartogràfic i Geològic de Catalunya (ICGC)

[13] Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

[14] Department of Earth and Environmental Sciences, Ludwig-Maximilian University of Munich, Munich, Germany

Declaration of Competing Interests: The authors acknowledge there are no conflicts of interest recorded.

Corresponding Author

Angelo Strollo email: strollo@gfz-potsdam.de

Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences

Telegrafenberg, 14473 Potsdam, Germany

Abstract

The European Integrated Data Archive (EIDA) is the infrastructure that provides access to the seismic waveform archives collected by European agencies. This distributed system is managed by Observatories and Research Facilities for European Seismology (ORFEUS). EIDA provides seamless access to seismic data from twelve data archives across Europe by means of standard services, exposing data on behalf of hundreds of network operators and research organizations. More than 12,000 stations from permanent and temporary networks equipped with seismometers, accelerometers, pressure sensors and other sensors are accessible through the EIDA federated services. A growing user base currently counting around 3000 unique users per year has been requesting data and using EIDA services. The EIDA system is designed to scale up to support additional new services, data types and nodes. Data holdings, services and user numbers have grown substantially since the establishment of EIDA in 2013. EIDA is currently active in developing suitable data management approaches for new emerging technologies (e.g. Distributed Acoustic Sensing) and challenges related to big datasets. This paper reviews the evolution of EIDA, the current data holdings and service portfolio and gives an outlook on the current developments and the future envisaged challenges.

Introduction

The rapid growth of seismology in Europe - from the ~30 openly available off-line stations in 1987 when the Observatories and Research Facilities for European Seismology (ORFEUS) was initiated, to the current situation with thousands of stations, many available in near real-time - resulted in ORFEUS evolving from a single centralized archive to a federated set of data centres. The concept of a centralized ORFEUS data centre thus changed through a number of European infrastructure projects (e.g. NERIES, NERA, SERA, EPOS-PP, EPOS-IP) towards an efficient federation of data centres (nodes) governed within ORFEUS, which is now known as the European Integrated Data Archive (EIDA) (see Data and Resources). The process, envisaged in 2011 was completed with the formal establishment of EIDA in 2013.

EIDA is the federated data and service infrastructure within ORFEUS aiming at secure archiving of seismic waveform data and metadata gathered by European seismic networks and research infrastructures, and provide the geoscience research communities seamless access to data. The EIDA archive is comprised of more than 12,000 stations with focus on the Euro-Mediterranean region (Figure 1), affiliated with both permanent and temporary networks. These networks are operated by the data suppliers, and include data from broad-band, short-period, strong motion, infrasound, and Ocean Bottom Seismometers (OBS) sensors. Table 1 provides an overview of the available stations by status, sensor type and accessibility. Data are disseminated via twelve EIDA nodes distributed across Europe in a seamless way through a suite of standardized web services. With the establishment of EIDA, ORFEUS significantly increased the availability of data from 12 TB in 2011 to 170 TB in 2013 and to nearly 600 TB in 2020 (Figure 2). The transition from a single data centre towards a distributed system produced not only technical challenges, but also required a new managerial structure. While the technical challenge was addressed by the usage of the ArcLink protocol (see Data and Resources), a new tailored governance structure was designed and implemented among the six initial founding nodes: ODC/KNMI: Orfeus Data Centre / Royal Netherlands Meteorological

Institute, De Bilt, Netherlands; GFZ: Helmholtz Centre Potsdam - German Research Centre for Geosciences, Germany; RESIF: Reseau Sismologique & Géodésique Français; ETHZ: Swiss Seismological Service, Zurich, Switzerland; INGV: Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy; BGR: Federal Institute for Geosciences and Natural Resources (BGR), Germany. In addition to the founding institutions, two data archives were added as 'secondary nodes': the Ludwig Maximilians Universität München (LMU), which is still part of EIDA, and the Institut de Physique du Globe de Paris (IPGP), for which data were later integrated into the RESIF EIDA node. A secondary node fully complies with the EIDA infrastructure in a technical sense, but does not contribute to the management level, where it is represented by a primary node (LMU is associated to BGR). Currently, EIDA comprises twelve nodes, with the following institutions joining as primary nodes over the years: the National Institute for Earth Physics (NIEP), Romania and the Kandilli Observatory and Earthquake Research Institute (KOERI), Turkey, in 2014; the National Observatory of Athens (NOA), Greece, in 2016; the University of Bergen (UiB) and NORSAR, Norway, jointly in 2019. The Instituto Cartográfico y Geológico de Catalunya (ICGC), Spain, entered as a secondary node in 2020 associated with ODC/KNMI.

Since its inception, the governing bodies of EIDA, the EIDA Management Board (EMB) and the EIDA Technical Committee (ETC), coordinate the day-to-day operations and define strategic developments. Through the EC-funded projects EPOS-IP and SERA, leveraging on other European projects (e.g. EUDAT, ENVRI+ and ENVRI-FAIR, EOSC-hub), and with the commitment of the individual EIDA nodes, EIDA has undergone major technical modernization in recent years. The initial technical architecture for data delivery based on ArcLink software has been progressively abandoned to increase robustness and scalability of the system as well as to rely on standard interfaces rather than specific software. As a replacement, EIDA has adopted the standard web services defined by the seismological community (see Data and Resources). EIDA data access today is based on a portfolio of web services which continue to be extended through collaborative work among the data centres coordinated within ORFEUS and in synergy with the European Plate Observing System (EPOS) developments (see Data and Resources). In this paper, we present the EIDA service infrastructure from standard services of the International Federation of Digital Seismograph Networks (FDSN) to external clients using these EIDA services through ad-hoc developed services to implement the distributed and interoperable system.

The EIDA service portfolio

EIDA provides access to seismic waveforms and associated quality information through the various nodes, as shown in Figure 3. Access to more than 12,000 stations from National Research Infrastructures across Europe is provided via a suite of standard FDSN web services and additional web services developed by the ETC. In the following subsections each EIDA service will be introduced briefly, see Data and Resources for documentation and usage about all services.

Fdsnws-station

The standard FDSN "station" web service operates at each EIDA node to expose metadata from the stations archived there. The service can provide StationXML (currently version 1.1) and text as standard output formats. The input parameters allow basic discovery of the available station inventory based on time in which the station was operative, geographical location, and any of the components of the NSLC codes (network, station, location, channel). The level of detail of the response can be specified at the moment of the query, from a simple list of networks, or stations, or streams (including all the details related to them), to StationXML format with all information available, including the response of each stream necessary to process the data. Most of the nodes use the implementation available in SeisComP (Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences and GEMPA GmbH, 2008), but as the interoperability is ensured via the use of standardized output, each data centre is free to choose their implementation.

Fdsnws-dataselect

The standard FDSN "dataselect" service is implemented by each EIDA node to retrieve waveform data in miniSEED format, the standard for seismic waveform data within the FDSN. This service handles all EIDA data requests. Its interface is almost identical to the fdsn-station web service, but without the capability to select by geographical location. In addition to that, this web service includes a method (queryauth) to request restricted data by authorized users. The authentication is based on the basic digest authentication implemented in all usual web servers. Users should be aware that it is common for EIDA implementations of this service to apply limits to the amount of data which can be requested. Thus, large requests should be split into sets of smaller requests. For those cases, there are some smart clients which can do this automatically (e.g. fdsnws2sds). All but one EIDA node are running the SeisComP implementation of fdsws-dataselect including authentication and authorization support.

WFCatalog

This quality control service (Trani et al., 2017) has been developed to meet user requests to be able to evaluate data quality before download and hence query for relevant data only. The web service provides detailed information on the contents of the waveform data in an archive. In particular the following features and quality parameters are provided: gaps, statistical values of background noise, availability, overlaps, quality flags. It may be used for quickly exploring metrics calculated on the waveforms before downloading the data, or by clients to fulfill user specific requirements. The API follows the style used by the FDSN web services with some specific additional features. The metrics are computed on fixed daily intervals (day boundaries), in case of gaps metrics are computed for each continuous data segment within the given day. The service has been developed by ODC/KNMI.

Routing service

The Routing Service (Quinteros, 2017) was designed to assist users and clients to locate data within a federated, decentralized collection of data centres such as EIDA. This service can be queried in order to locate the data. Some smart clients (fdsnws_fetch, ObsPy routing client) are also provided to offer the user an integrated view of the entire EIDA, hiding the complexity of its internal structure. This service was developed starting from the former concept of routing table used previously in EIDA when federation was first implemented using ArcLink. It allows scalability both in terms of nodes and services. The current EIDA routing service is not only used by the users to locate specific data but is also used to synchronize the routes within the federation, replacing the old synchronization mechanism that involved merging and replicating the EIDA inventory at all nodes. Today, only the routes are synchronized and merged such that all nodes know about all routes. The routing service also allows definition of primary (authoritative) routes alongside secondary routes. If more than one node attempts to declare conflicting primary routes for data, this will be detected and moderated. The service harvests the routes (datasets) declared by all nodes to expose them in the official EIDA Routing Service running at ODC, where client tools expect to find it. The approach used by the Routing Service to cope with the declaration of services, datasets and priorities in case of multiple copies, has been adopted by FDSN in its last FDSN Plenary Meeting (Montreal, 2019). There, a standard metadata schema was approved to declare data centres and their available datasets. Thus, all information available in the EIDA Routing Service can be harvested by FDSN to provide a detailed declaration of the services and data holdings available at the nodes (see Data and Resources). The service has been developed by GFZ.

Authentication service

Although only a minor part of EIDA data are embargoed, a new authentication and authorization infrastructure (AAI) for the federated system was needed in order to deal with these restrictions, to get better statistics regarding data usage, and also to possibly expand services in the future for authenticated users. In order to have a system that is easy to maintain and that is scalable with services and data in a federated context, authentication and authorization have been decoupled. The service redirects the user to its home institution (if it is affiliated to eduGAIN) to sign in and, in the case of a successful sign in, it provides a digitally signed token in which all available attributes are present. While this works very well for authentication, it may be insufficient to decide if access to restricted data should be granted. This is because the home institution may not send enough attributes to our system due to different regulations. To overcome this issue, we introduced a B2ACCESS instance working as a proxy to eduGAIN. B2ACCESS allows also users from institutions not affiliated to eduGAIN to sign in with alternative accounts (e.g. ORCID, GitHub) or sign up for a new account (see Data and Resources). Through B2CCESS, Principal Investigators (PIs) or data centre operators can define the attributes needed to give permissions to users, decoupling what is needed for authentication (provided by eduGAIN) from what is needed by the Authorization (defined internally by the data centres). The token can then be presented to any EIDA service (e.g. fdsnws-dataselect). Services can check the token integrity and, based on the information stored there, grant access to temporary project data (e.g. AlpArray Seismic Network, Hetényi et al., 2018) without exchanging and updating the user access control list at all nodes. Clients like WebDC3, fdsnws_fetch, and the ObsPy routing client, as well as an extended version of fdsnws-dataselect running at EIDA nodes, support the use of tokens. The service has been developed by GFZ.

Federator and other clients

Federator

The EIDA Federator can automatically route and retrieve requests federated between EIDA nodes. It provides a single service entry point for the entire EIDA holdings for the fdsnws-station and -dataselect services, and the EIDA WFCatalog service. This prevents users from having to query the routing service before making data requests to the individual EIDA nodes. The federator currently does not support authentication and hence cannot be used to download restricted datasets. Specifications are compliant to the federated web services at the end points: (fdsnws-station, fdsnws-dataselect, WFCatalog). The service has been developed and is operated by ETHZ.

Fdsnws_fetch

Fdsnws_fetch is a distributed data request tool that is based on FDSN web services and the EIDA routing service, developed to ease the transition of former arclink_fetch users to FDSN web services. It supports tokens generated by the EIDA authentication service and provides a proper citation for the data requested based on the FDSN citation service. This smart client has been developed by GFZ.

ObsPy

An additional smart client has been developed in collaboration with the ObsPy community (Beyreuther et al., 2010), namely the routing client which is capable of using the routing service and the EIDA authentication service. Similar to fdsnws_fetch the routing client is requesting data directly from the end points after liaising with the routing and authentication central services (Figure 3). This smart client is widely used to access EIDA data, as a large number of EIDA users use ObsPy / Python for download and data processing.

EPOS-ICS

The EPOS-ICS portal is successfully integrating the EIDA services, and the EIDA services are indeed part of the EPOS Seismological Thematic Core Service (EPOS-S TCS). Therefore, data from EIDA is made available to the geoscience community along with other types of data through an interoperable platform for data discovery, access, and processing. This includes also the concept of virtual networks which is used for example by the AlpArray community, and the EPOS Near Fault Observatories (NFO) to collect related stations together.

Interactive access

The EIDA portal also provides browser-based interactive access using WebDC3 (Bianchi et al., 2015). WebDC3 is a web interface to SeisComP standard seismological services allowing users to conveniently discover seismic stations distributed from EIDA data centres, explore events in a number of 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. The web interface supports tokens generated by the EIDA authentication service.

User feedback and documentation

In 2019, triggered by the AlpArray user community, the EIDA developers opened an EIDA User Feedback Repository hosted on GitHub. This has become the preferred way of reporting a wide variety of issues to the EIDA Maintenance Team (technical difficulties, questions, suggestions).

Extensive documentation for the users is also provided and kept up to date next to the services access pages at the ORFEUS portal including specific documentation for data centre operators.

Data management and dissemination

EIDA data policies are coordinated among the various EIDA nodes within ORFEUS. When becoming an EIDA node, a new data centre must commit to an open data policy, is expected to demonstrate a minimum service availability (95%), ensuring in house at least one redundant copy of the data for hot backup and optional cold backup off site (in place at some nodes). Ingestion of data and metadata at each node is performed according to its own procedures with all nodes requested to carefully check consistency of metadata with data suppliers. Some of the EIDA services are synchronized daily through the nodes (e.g. routing service and logging) and all are monitored via their own or federated tools. At all data centres a dedicated technical person is available to provide response to arising issues, in general within one working day. Through the use of standard services, metadata and data formats, the long term commitment of all EIDA nodes to curate data with a good level of data FAIRness (Wilkinson et al., 2016) is also achieved.

EIDA ensures that seismic networks within its archives have a DOI (Digital Object Identifier), following the FDSN guidelines (Evans et al., 2015) and most EIDA nodes can mint a DOI for the networks they host. EIDA promotes rich metadata, including standard licences where possible (e.g. Creative Commons CC BY 4.0 or similar). DOIs are also integrated in the seismic station metadata (FDSN StationXML) to formally establish the link with the DOI metadata that can be automatically harvested for a specific seismic network. EIDA requires the users to provide proper reference to the data suppliers. This can be done by citing the seismic networks with their associated DOIs, or in rare cases when these are unavailable, by using network name and/or FDSN network code.

Taking advantage of the development in the SERA project (see Data and Resources) the EIDA group has been working on metadata challenges and proposed solutions to also integrate other types of data than those historically present in the EIDA nodes (e.g. OBS, Infrastructure monitoring, Infrasound). Following these developed guidelines data and metadata from the French and the German OBS pools, collected within the AlpArray project (AlpArray Seismic Network, 2015) have been archived for the first time in a consistent way at more than one node with common pre-processing and metadata creation procedures. In the process of improving workflows, also for network operators and data suppliers, EIDA developers participate in the development of the new FDSN documentation for StationXML, partially sponsored also by ORFEUS.

Figure 4 shows the data disseminated by EIDA to a base of more than 3000 annual unique users with increasing requests, up to 180 TB/year in 2019. Since 2020 data are only distributed via web services as shown in Figure 4. Web services have rapidly gained popularity starting from 2016, enabling the shutdown of ArcLink in 2019.

Discussion and Conclusions

With the adopted governance and technical setup, the EIDA system has demonstrated that federated approaches, although more complex and difficult to maintain, are a valid solution to serve users of seismological data when resources are distributed and for political and financial reasons not available in a single institution. With the growing demand for large volumes of data, distributed archives have become a very attractive solution to minimize failures related to single access points. The fundamentals of the present EIDA system is a modular scalable infrastructure based on standard interfaces. Development of a logging system fully compliant with the European General Data Protection Regulation (GDPR) is in progress alongside with a new interactive portal to access data making use also of the quality metrics to let users assemble tailored data sets. The development is carried out in a collaborative and coordinated framework among the technical group with guidance from the management board.

EIDA is not only a technical development infrastructure but also a coordination group within ORFEUS that can provide informed proposals and opinions within the FDSN reducing fragmentation. Recent successful examples of coordinated efforts within the FDSN are the guidelines for DOI, jointly prepared and updated; the registry of data centres at the FSDN and the integration of the routing information; the StationXML documentation and various ongoing discussions that will evolve in proposals for FDSN standards (e.g. Large-N and Distributed Acoustic Sensing (DAS) data management, new authentication method following an approach similar to the one in place for EIDA).

Following valuable feedback from the User Advisory Group (UAG) of ORFEUS, quality metrics and data and metadata harmonization across EIDA are currently in focus. For the former, EIDA is exploring the possibility of aligning quality metrics from WFCatalog (Trani et al, 2017) with those produced by MUSTANG (Casey et al., 2018) to achieve standardization within FDSN. Harmonization in terms of enhancing best practices within EIDA for data suppliers including temporary deployments is among the priorities of the group as well.

EIDA continues to grow not only in terms of archive volume, but also in other areas: the number of nodes is increasing; new datasets are becoming relevant for seismologists, such as DAS or cheap seismic sensors (Quinteros et al., 2021); users demand rapid access to massive data volumes. These issues will continue to challenge EIDA in the near future. In order to prepare for these challenges EIDA is collaborating with IRIS, leveraging on EU projects (see Data and Resources), to prepare data management concepts for emerging technologies such as DAS, massive numbers of cheap sensors and microsensors for the internet of things (IoT). Whatever the challenges will be, EIDA is ready to engage in new developments taking advantage of the distributed setup and thus the sustainable framework.

Data and Resources

We list below all online resources used in this work. They were last accessed in October 2020.

- EIDA and ORFEUS: http://www.orfeus-eu.org/data/eida/.
- ArcLink protocol: https://www.seiscomp.de/seiscomp3/doc/applications/arclink.html.

- EPOS: https://www.epos-eu.org/
- FDSN web service specifications: https://www.fdsn.org/webservices/.
- Documentation about usage of fdsnws-station, fdsnws-dataselect, routing, WFCatalog, federator: http://www.orfeus-eu.org/data/eida/webservices/.
- Authentication service: https://geofon.gfz-potsdam.de/eas/.
- FDSN's data centre registry: http://www.fdsn.org/datacenters/
- eduGAIN: https://edugain.org/.
- B2ACCESS: https://eudat.eu/services/b2access.
- Federator services (station, dataselect and WFCatalog): http://eida-federator.ethz.ch/fdsnws/.
- Fdsnws_fetch and fdsnws2sds: https://geofon.gfz-potsdam.de/software/fdsnws_scripts/.
- The ORFEUS WedDC3 portal can be accessed at http://orfeus-eu.org/webdc3/ where also additional documentation is available.
- The Issue tracker of the user feedback repository is accessible at: https://github.com/EIDA/userfeedback/issues.
- Documentation for data centre operators: https://orfeus-eu.readthedocs.io/en/latest/.
- ObsPy's routing client: https://docs.obspy.org/packages/autogen/obspy.clients.fdsn.routing.routing_client.html.
- The EPOS ICS portal for integrated access to earth science data is accessible at: https://www.ics-c.epos-eu.org/.

- The SERA project deliverable “Report on metadata challenges and proposed solutions” can be downloaded at: http://www.sera-eu.org/export/sites/sera/home/.galleries/Deliverables/SERA_D4.2_Metadata-challenges-and-proposed-solutions.pdf.
- Data distributed via EIDA services are provided by many data suppliers. The complete list of seismic networks (110 permanent and 206 temporary networks) can be found at: http://www.orfeus-eu.org/data/eida/networks/.
- EOSC-Pillar web site: https://www.eosc-pillar.eu/.
- RISE web site: http://www.rise-eu.org/home.

Acknowledgments

EIDA has been built over a decade, involving contributions from many people over time. First of all, EIDA would like to thank the networks who entrust data to EIDA nodes, and the scientific users for their trust and continuous feedback. We thank the various people involved in the governance of ORFEUS and in particular its Secretary General Carlo Cauzzi and the EPOS Seismology Secretary Florian Haslinger for their coordination efforts and continuous support. We would also like to thank the many previous participants in EIDA and the EIDA nodes, without whom the present data distribution system would not have existed. Seismological data distribution in Europe started well before EIDA; all of the EIDA team is grateful for the legacy of the initial pioneers of the open data distribution in Europe, and the initiators of the integrated archive Winfried Hanka and Domenico Giardini. EIDA would like to pay tribute to Torild van Eck who tirelessly promoted opening data from the hundreds of seismic networks in Europe and adjoining areas. EIDA has benefited from financial support from various EU projects (NERIES, NERA, EPOS-PP, EPOS-IP, SERA, EUDAT, EOSC-hub, EOSC-Pillar Grant Number 857650, RISE Grant Number 821115 and EPOS-SP Grant Number 871121). The EIDA nodes have received financial support from institutional funding and in some cases national projects. Comments and suggestions from the anonymous reviewers and the editor are also acknowledged.

References

AlpArray Seismic Network (2015): AlpArray Seismic Network (AASN) temporary component. AlpArray Working Group. Other/Seismic Network., doi: 10.12686/alparray/z3_2015

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

Bianchi, Marcelo; Evans, Peter L.; Heinloo, Andres; Quinteros, Javier (2015): WebDC3 Web Interface. GFZ Data Services. Doi: 10.5880/GFZ.2.4/2016.001

Casey, R., M. E. Templeton, G. Sharer, L. Keyson, B. R. Weertman, T. Ahern (2018). Assuring the Quality of IRIS Data with MUSTANG, Seismol. Res. Lett., doi: 10.1785/0220170191.

Evans, P., Strollo, A., Clark, A., Ahern, T., Newman, R., Clinton, J. F., Pedersen, H., Pequegnat, C. (2015): Why Seismic Networks Need Digital Object Identifiers. - Eos, Transactions American Geophysical Union, 96., doi: 10.1029/2015EO036971.

Hetényi, G., Molinari, I., Clinton, J. et al. The AlpArray Seismic Network (2018): A Large-Scale European Experiment to Image the Alpine Orogen. Surv Geophys 39, 1009–1033., doi: 10.1007/s10712-018-9472-4

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

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., Carter, J. A., Schaeffer, J., Trabant, C., Pedersen, H. A. (2021). Exploring Approaches for Large Data in Seismology: User and Data Repository Perspectives. Seismological Research Letters 2021; doi: https://doi.org/10.1785/0220200390

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

Wilkinson, M., Dumontier, M., Aalbersberg, I. *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 authors

Angelo Strollo - strollo@gfz-potsdam.de - ORCID 0000-0001-9602-6077 - Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, German

Didem Cambaz - didem.samut@boun.edu.tr - Bogazici University, Kandilli Observatory and Earthquake Research Institute, Istanbul, Turkey

John Clinton - jclinton@sed.ethz.ch - ORCID: 0000-0001-8626-2703 - Swiss Seismological Service, ETH Zurich, Zurich, Switzerland

Peter Danecek - peter.danecek@ingv.it Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

Christos Evangelidis - cevan@noa.gr - ORCID: 0000-0001-8733-8984 - National Observatory of Athens, Institute of Geodynamics, Athens, Greece

Alexandru Marmureanu - marmura@infp.ro National Institute for Earth Physics (NIEP), Romania

Lars Ottemöller - lars.ottemoller@uib.no - ORCID: 0000-0003-3170-1048 - Department of Earth Science, University of Bergen, Norway

Helle A. Pedersen - helle.pedersen@univ-grenoble-alpes.fr - ORCID: 0000-0003-2936-8047 - Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France

Reinoud Sleeman - sleeman@knmi.nl Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

Klaus Stammler - klaus.stammler@bgr.de Federal Institute for Geosciences and Natural Resources (BGR), Germany

Daniel Armbruster - daniel.armbruster@sed.ethz.ch Swiss Seismological Service, ETH Zurich, Zurich, Switzerland

Jarek Bienkowski - jarek.bienkowski@knmi.nl Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

Kostas Boukouras - kbouk@noa.gr National Observatory of Athens, Institute of Geodynamics, Athens, Greece

Peter L. Evans - pevans@gfz-potsdam.de Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Germany

Massimo Fares - massimo.fares@ingv.it - ORCID: 0000-0002-1399-0379 - Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

Cristian Neagoe - cristian.neagoe@infp.ro - ORCID: 0000-0002-2690-5794 - National Institute for Earth Physics (NIEP), Romania

Stefan Heimers - heimers@sed.ethz.ch Swiss Seismological Service, ETH Zurich, Zurich, Switzerland

Andres Heinloo - andres@gfz-potsdam.de Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany

Mathias Hoffmann - mathias.hoffmann@bgr.de Federal Institute for Geosciences and Natural Resources (BGR), Germany

Philippe Kaestli - kaestli@sed.ethz.ch - ORCID: 0000-0001-8667-8549 - Swiss Seismological Service, ETH Zurich, Zurich, Switzerland

Valentino Lauciani - valentino.lauciani@ingv.it - ORCID: 0000-0002-9672-7428 - Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

Jan Michalek - jan.michalek@uib.no - ORCID: 0000-0002-8057-7541 - Department of Earth Science, University of Bergen, Norway

Erich Odon Muhire - erichodon.muhire@bgr.de Federal Institute for Geosciences and Natural Resources (BGR), Germany

Mehmet Ozer - mehmet.ozer@boun.edu.tr Bogazici University, Kandilli Observatory and Earthquake Research Institute, Istanbul, Turkey

Lucian Palangeanu - lucian.palangeanu@yahoo.com National Institute for Earth Physics (NIEP), Romania

Constanza Pardo - pardo@ipgp.fr - ORCID: 0000-0003-1601-8718 - Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France

Javier Quinteros - javier@gfz-potsdam.de - ORCID: 0000-0001-9993-4003 - Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Germany

Matteo Quintiliani - matteo.quintiliani@ingv.it - ORCID: 0000-0002-9453-4155 - Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Nazionale Terremoti, Rome, Italy

Jose Antonio Jara Salvador - JoseAntonio.Jara@icgc.cat Institut Cartogràfic i Geològic de Catalunya (ICGC)

Jonathan Schaeffer - jonathan.schaeffer@univ-grenoble-alpes.fr - ORCID: 0000-0002-2447-7394 - Univ. Grenoble Alpes, Irstea, CNRS, IRD, Météo France, OSUG, 38000 Grenoble, France

Antje Schloemer - antje.schloemer@lmu.de Department of Earth and Environmental Sciences, Ludwig-Maximilian University of Munich, Munich, Germany

Nikolaos Triantafyllis - triantafyl@noa.gr - ORCID: 0000-0002-4011-0971 - National Observatory of Athens, Institute of Geodynamics, Athens, Greece

Tables

Table 1: Available stations via EIDA by status, sensor type and accessibility (updated to October 2020).

Status	**Operational**	**Dismantled**	**Total**
	4,155	8,163	12,318
Sensor type	**Seismometer**	**Accelerometer**	**Pressuremeter**
	11,303	1,750	166
Accessibility	**Open**	**Embargoed**	
	9,572	2,746	

List of figure captions

Figure 1

Geographical distribution of the 12,000+ stations available from EIDA: a) operational (green) versus dismantled stations (orange); b) permanent (blue) versus temporary stations (red); c) by sensor type: seismometers in green [velocity], strong motions in red [acceleration] and OBS or infrasound in blue [pressure]; d) zoom on central Europe with strong motion only (red). Light blue squares indicate the EIDA nodes.

Figure 2

ORFEUS data holdings evolution since 2011. Numbers for 2011 reflect data holdings available at the ORFEUS Data Centre (ODC). The cumulative EIDA volume, including also ODC has been added also for 2012 although EIDA was formally established only in 2013. Red text indicates data nodes present from the start and (within the histogram) new nodes.

Figure 3

Schematic view of the EIDA infrastructure (right) and user's workflow (left). In the central ring are represented the Data Suppliers providing network/stations distributed via the EIDA nodes. The outer ring contains EIDA nodes and the distributed services. In the left part three possible user's workflow are represented: a) the user sends a data request via smart client that will get routes from the central routing service then contact the necessary nodes and provide data back to the user (dashed line); b) the user requests data directly to the the nodes (solid line); c) the user requests data via a centralized service that will act as a proxy requesting the actual data to the nodes and provide them to the user (dotted line).

Figure 4

Figure 4: Data distribution through the years since the formal establishment of EIDA (2013). In light grey the yearly data volume distributed via fdsnws and ArcLink. In red the yearly volume counting only fdsnws, for 2020 data are incomplete and only fdsnws since ArcLink services were stopped at all nodes in December 2019. Note that real-time data distribution from each data node is significant, but is not included as this data distribution is not part of the EIDA service catalogue.

Figures

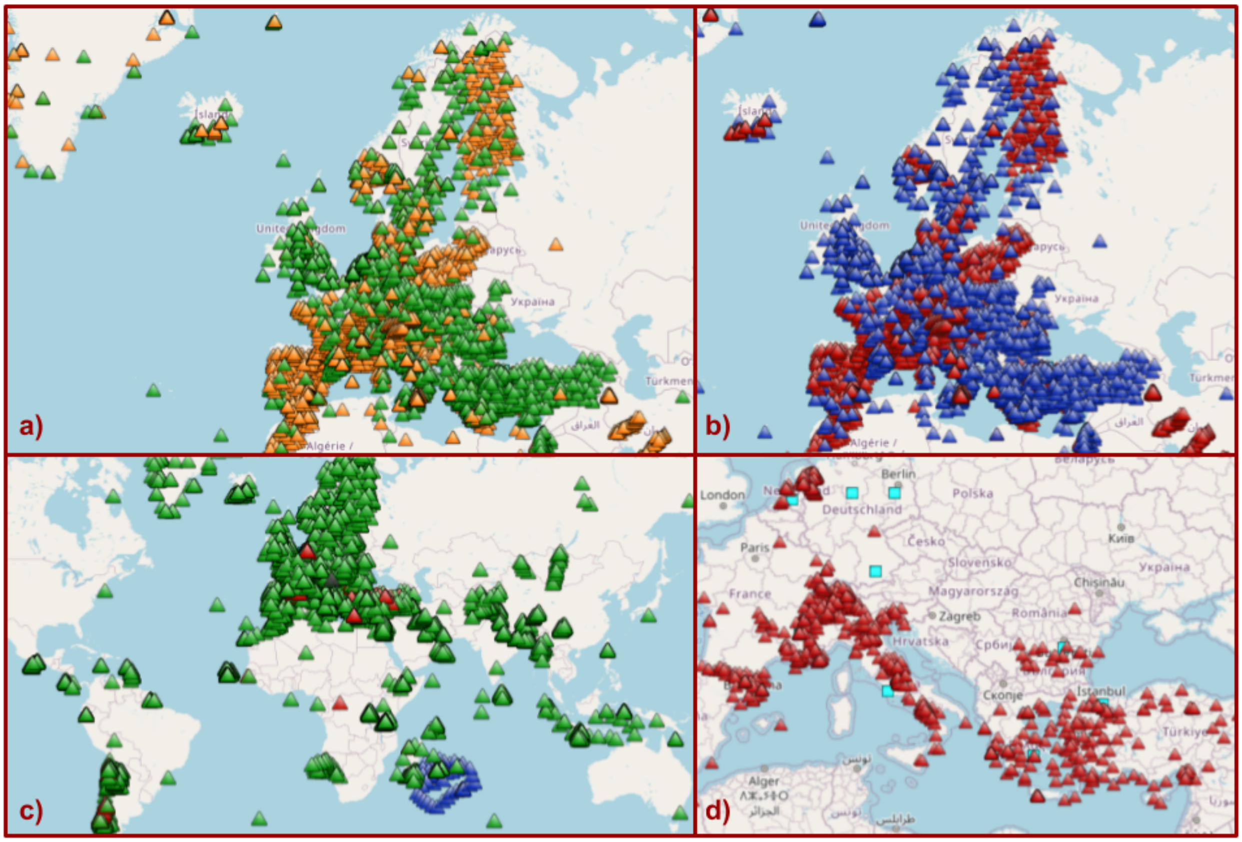

Figure 1: Geographical distribution of the 12,000+ stations available from EIDA: a) operational (green) versus dismantled stations (orange); b) permanent (blue) versus temporary stations (red); c) by sensor type: seismometers in green [velocity], strong motions in red [acceleration] and OBS or infrasound in blue [pressure]; d) zoom on central Europe with strong motion only (red). Light blue squares indicate the EIDA nodes.

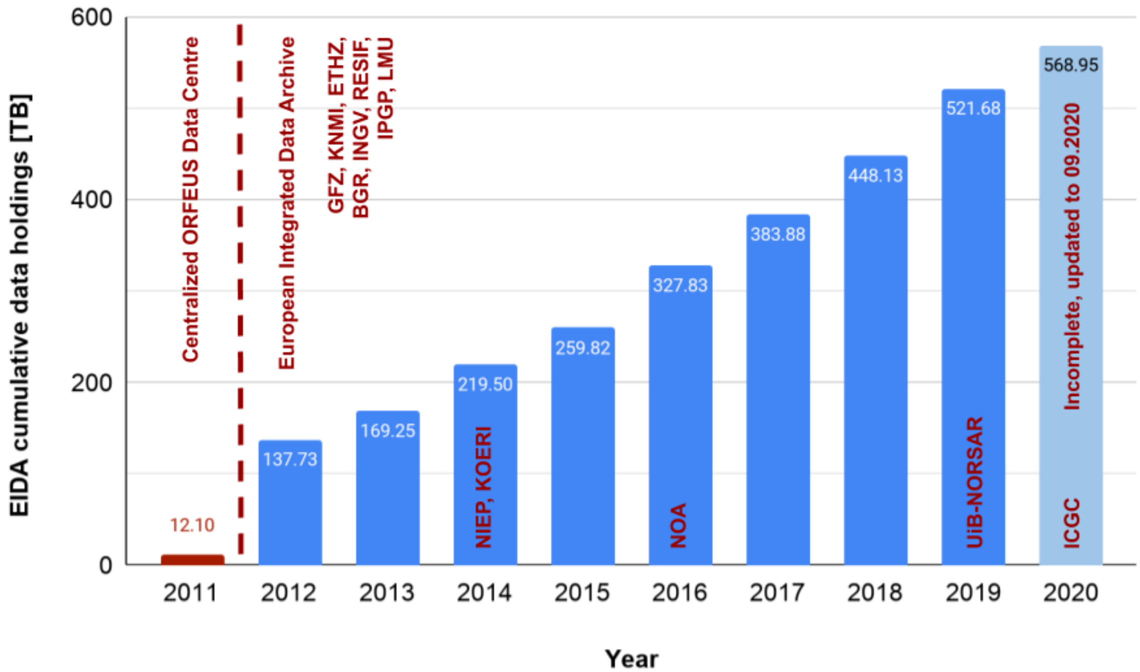

Figure 2: ORFEUS data holdings evolution since 2011. Numbers for 2011 reflect data holdings available at the ORFEUS Data Centre (ODC). The cumulative EIDA volume, including also ODC has been added also for 2012 although EIDA was formally established only in 2013. Red text indicates data nodes present from the start and (within the histogram) new nodes.

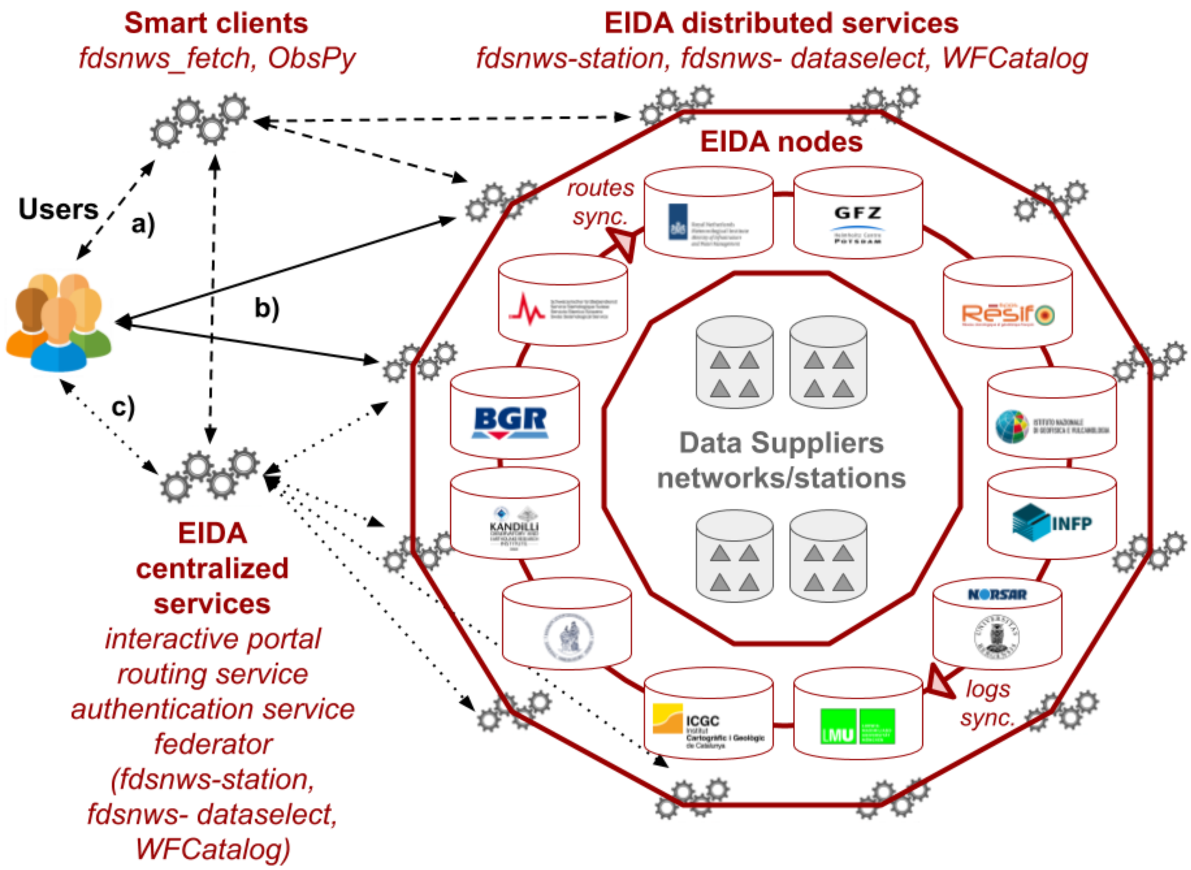

Figure 3: Schematic view of the EIDA infrastructure (right) and user's workflow (left). In the central ring are represented the Data Suppliers providing network/stations distributed via the EIDA nodes. The outer ring contains EIDA nodes and the distributed services. In the left part three possible user's workflow are represented: a) the user sends a data request via smart client that will get routes from the central routing service then contact the necessary nodes and provide data back to the user (dashed line); b) the user requests data directly to the nodes (solid line); c) the user requests data via a centralized service that will act as a proxy requesting the actual data to the nodes and provide them to the user (dotted line).

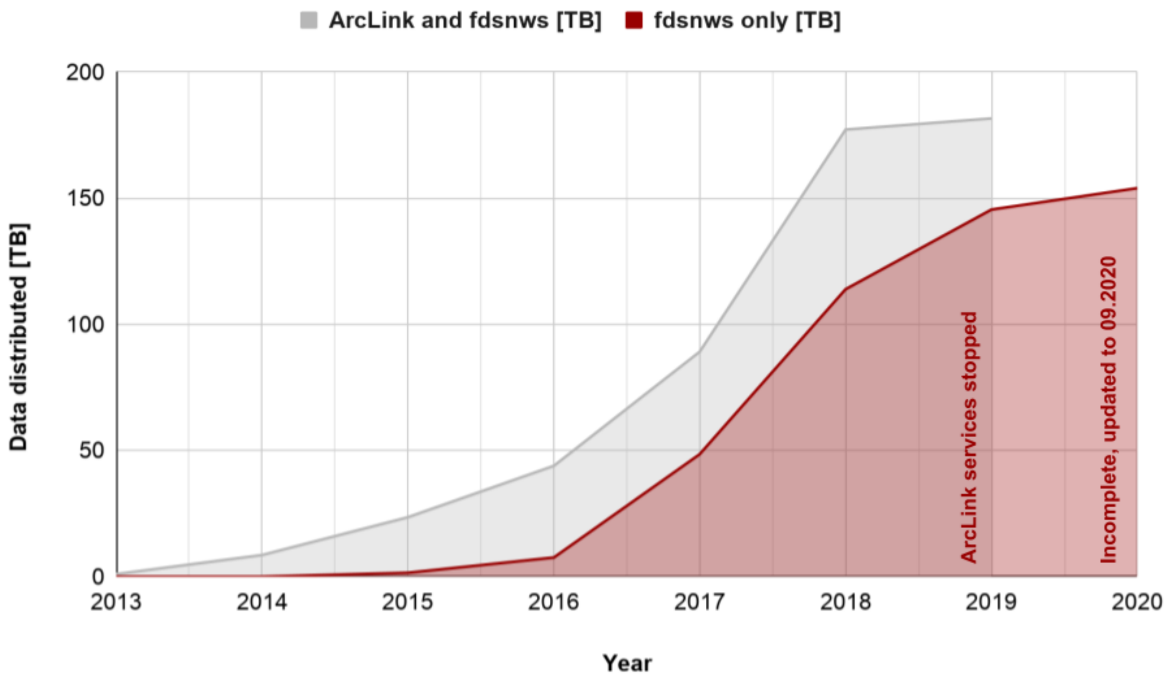

Figure 4: Data distribution through the years since the formal establishment of EIDA (2013). In light grey the yearly data volume distributed via fdsnws and ArcLink. In red the yearly volume counting only fdsnws, for 2020 data are incomplete and only fdsnws since ArcLink services were stopped at all nodes in December 2019. Note that real-time data distribution from each data node is significant, but is not included as this data distribution is not part of the EIDA service catalogue.