SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D20.1 Report on access statistics and service provision of VA1-VA5

(PART 1/EMSC)

Work package	WP18 - VA1: Virtual Access to seismological products and information at EMSC
Lead	EMSC
Authors	Matthieu Landès, EMSC; Julien Roch, EMSC; Rémy Bossu, EMSC
Reviewers	R. Bossu (EMSC), K. Saleh (ETH)
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	29.02.2020
Submission date	30.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D20.1_Report_Virtual_Access_statistics.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

The aim of this document is to describe the data quality and the traffic monitoring of the services provided by the European-Mediterranean Seismological Centre (EMSC). This document follows the same structure of the previous deliverable D18.1. Service descriptions have been kept for the self-consistency of the document. However, all statistics has been updated with traffic measurements since September 2018.

EMSC is a key actor for global earthquake information and is involve in the seismological community, in scientific European projects and in communication media for the general public. Since the initial SERA proposal, EMSC is constantly evolving. The volume of collected data from seismological institutes and eyewitnesses is increasing. Moreover, the same trend is observed on the traffic monitoring of the different EMSC services. On the websites, on Twitter, on the Seismic Portal or on the mobile application LastQuake, EMSC gains popularity among seismologists and among the general public.

This effort for providing good quality data and information imposes more and more constraints on the EMSC core real time system. It has to be faster and more flexible while maintaining its reliability. This work is now in progress.

Statistic overview since September 2018

Since the first period of the SERA project, the EMSC continues to gain popularity and to be well identified with the general public. In general, all traffic measurements are increasing: the number of users of our mobile application LastQuake, the web and mobile traffic, followers of our social media accounts…

In addition to this trend, the seismic sequence in Albania started in November 2019 has had a huge impact on EMSC. To such an extent that some of our services were saturated some minutes during the main shock. This event lasted almost three weeks, it brokes many records of attendance and brings upward the statistics of the end of 2019.

1 Introduction

This chapter gives an overview of the European-Mediterranean Seismological Centre (EMSC) and describes the work done since September 2018 within the SERA project.

1.1 Overview of EMSC

EMSC is one of the very top global earthquake information centres. All activities are closely coordinated with EMSC members (85 institutes and observatories in 56 countries) as well as with US Geological Survey and with the International Association for Seismology and Physics of the Earth's Interior (IASPEI). Finally, EMSC is one of the pillars of the seismological services of the ESFRI[1] research infrastructure EPOS (European Plate Observing System).

EMSC provides real time earthquake information and earthquake products ranging from authoritative locations, moment tensors, global macroseismic data, to information related to earthquake's impact (qualitative impact estimates, mapping of the felt area, rapid detection of felt earthquakes, geo-located picks). EMSC also hosts the RESORCE[2] database, a reference database for specific Ground Motion Prediction Equations (GMPE) studies as well as a Quake Catcher Network (QCN) server to ease and favour the developments of citizen operated networks in the Euro-Med region.

The EMSC infrastructure can be divided in a collection part and a dissemination part (Figure 1).

- The **collection part** collates seismological data collected from national institutes and data collected from eyewitnesses that share their experiences through felt reports, comments or pictures.
- The **dissemination part** comprises EMSC classic mobile and desktop websites, the Seismic Portal for accessing and visualizing seismic data, accounts on social networks (Twitter, Facebook, Telegram) and a mobile application "LastQuake".

[1] *European Strategy Forum on Research Infrastructures*
[2] Reference database for Seismic grOund-motion pRediction in Europe

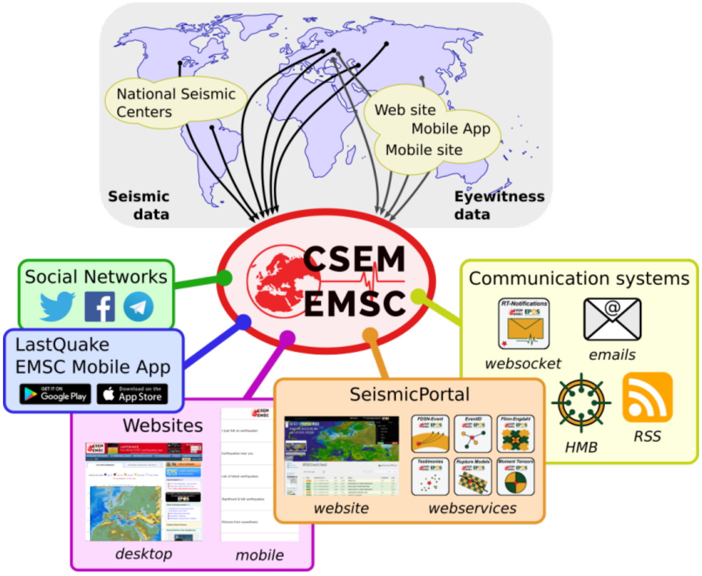


Figure 1: Overview of the main EMSC services

1.2 Description of work

1.2.1 Strengthening the EMSC communication system

The communication system of EMSC is one of the key components to extend and strengthen the EMSC services. For many years, for **data exchange with contributors**, we use emails and the messaging system PDL[3] developed by the USGS. Within the European project EPOS, EMSC has added the new HMB messaging system developed by GFZ that works on the TCP port 80**. On the user side**, the EMSC gives access to seismological data via 6 web services through the Seismic Portal. The FDSN-event exists since 2014 but the 5 others are new and have been released in the middle of 2017. In addition, the EMSC communicates real time information with a Twitter quakebot. Messages sent to Twitter are constantly evolving and participate to the increasing popularity of EMSC.

Since September 2018, a second front server has been setup for the Seismic Portal website and webservice. Moreover data contributors have been contacted again individually to renew the contacts and verify available seismic data.

1.2.2 Accessibility for eyewitnesses and seismologists

The initial proposal to separate website for eyewitnesses from the general one has evolved. We have noticed that a higher share of eyewitness reports are done through our mobile app LQ and the mobile website (basically: through smartphones). Although LastQuake is constantly developed, **our mobile website needs to be upgraded** and it's becoming a priority. This work has been initiated at the end of 2018. In addition we plan to move more scientific contents to the Seismic Portal. Moreover, on the hardware side, **the front end server of all EMSC services has been upgraded** thanks to the funding of the CEA that host the EMSC. This server is well designed to face huge traffic peak of user during seismic crisis without any slowdown.

[3] Production Distribution Layer developed by USGS, https://github.com/usgs/pdl

Since September 2018, the work on the mobile website has been initiated. We chose to keep the same design as the one used in LastQuake. A prototype is online and is tested internally. Our goal is to publish it as soon as possible. Moreover a lot of work has been done on the hardware side. Thanks to the CEA funding, EMSC informatics infrastructure is now based on virtual machines hosted on Dell FX2 servers. This system makes it easier to set up new web servers to have a better traffic distribution between front servers.

Concerning the Quake Catcher Network (QCN), the EMSC has developed a QCN server available in 2016. The website is accessible here: http://qcn.emsc-csem.org/ and it allows users to visualize and query QCN data (Figure 2). With the design of the sensors, the recorded waveforms are very sparse and with the lack of funding, this project is less popular and its future is compromised. On the topic of citizen seismology sensors, we are testing the *Raspberryshake* sensor-digitizer developed by OSOP[4].

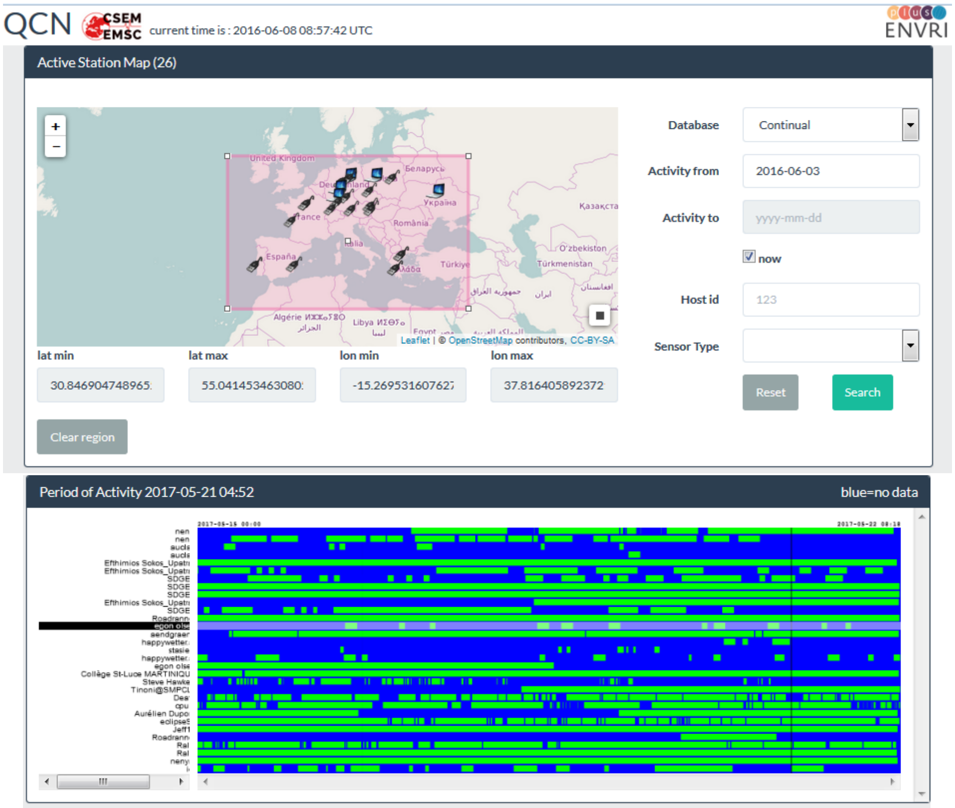


Figure 2: Graphical user interface of the EMSC QCN server. The user has the possibility to search for sensors (top) and to visualize data availability.

Within the EPOS project, the **Seismic Portal** has evolved and now contains new services that give access to felt reports, moment tensors, source models, mapping of event identifiers and Flinn-Engdahl region names. For each dataset, a graphic user interface and a web service are available.

1.2.3 Performance and reliability of the EMSC real time system

In the effort to strengthen its current services, the EMSC has begun an upgrade of its real time system in the beginning of 2018. This system is a legacy of 15 years of continuous developments and keeps

[4] http://www.osop.com.pa/

historical informatics choices. The actual EMSC real time system has the constraints for fast data processing, for fast communication and for more flexibility for new developments.

Since September 2018, two main actions have been done. First, our production servers where old and thanks to the founding of the CEA we have migrate most of them on virtual machines installed on new servers. Second, we have continued the refactoring of the core system. The code is now versioned and we have installed this version on our backup server at the Instituto Geografico Nacional in Madrid.

2 EMSC collection system

The EMSC **collection system** collates seismological data and interfaces with global earthquake eyewitnesses in order to massively crowdsource testimonies, comments and geo-located pictures.

Although this section keep the same structure as the D18.1 document, satistics and measurements are takes into account the last period of the SERA project which begin after September 2018.

2.1 Seismological data

Seismic data are collected in real time from 96 observatories (in 2019) from around the world. On average, we collect 15k seismic origins per month and 230 moment tensor solutions per month representing information for at least 4000 earthquakes monthly from all over the world (Figure 3). This set of data is the foundation of all EMSC dissemination services and is widely used by the seismological community.

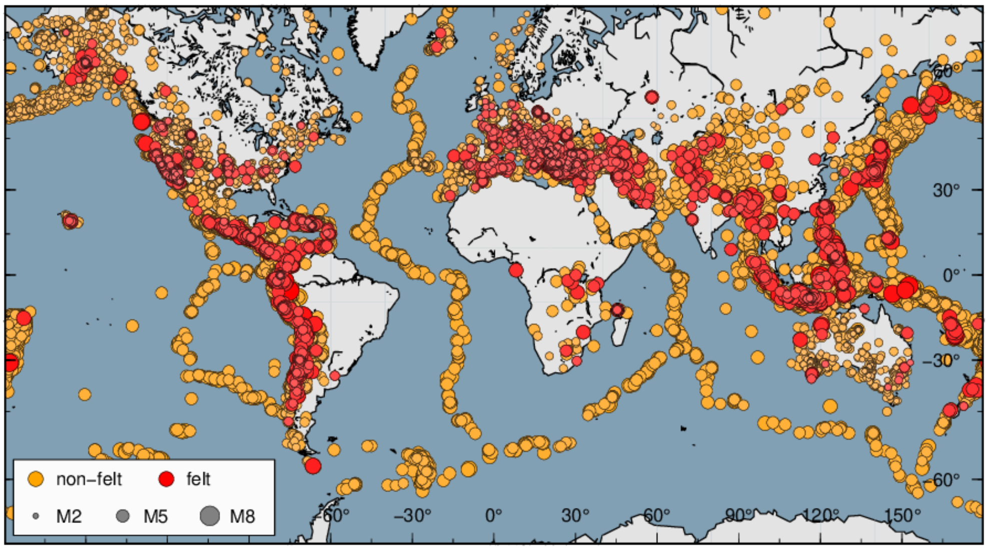


Figure 3: Earthquakes collected by the EMSC between September 2018 and end of December 2019. During this period, we have collected 80174 earthquakes among which 3593 are considered as felt earthquake.

The overall quality of EMSC data relies on the quality of contributors and in 2019 we have updated contact of almost all contributors. It’s an ongoing process that keeps active collaboration.

2.2 Data from eyewitness community

Eyewitness data are collected through our websites and our mobile app. People who feel an earthquake and want to share their experience have the choice to evaluate the level of shaking with thumbnails (felt reports), to write a comment or to send pictures. The following data was collected between 1st September 2018 and December 2019.

The map of all felt reports (Figure 4) shows a world wide distribution. EMSC covers almost all populated seismic regions even though in North Africa and Eastern Asia a lack of popularity can be observed. In any case, this world wide distribution gives us confidence in our ability to engage population wherever damaging earthquakes may strike.

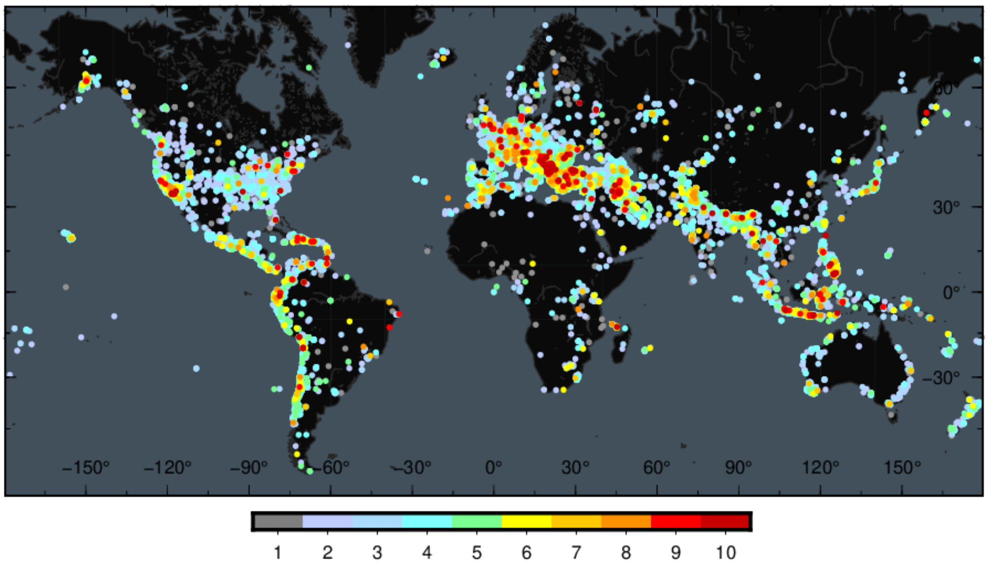


Figure 4: Distribution of the 249k felt reports collected by EMSC between September 2018 and December 2019. The color scale represents the level of shaking reported by eyewitnesses.

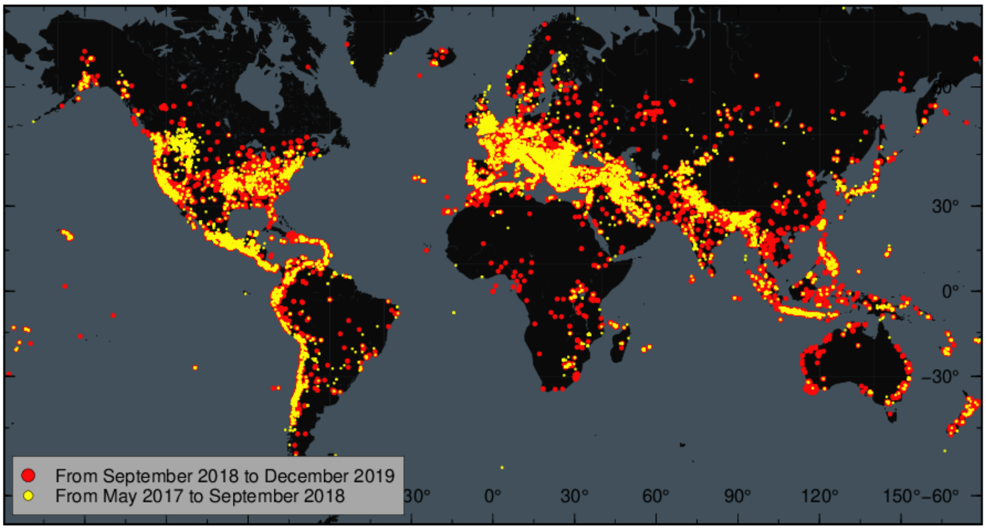


Figure 5: Comparison of felt report distribution between the period between September 2018 and December 2019, and the previous between May 2017 and September 2018.

Felt reports is our main indicator measuring EMSC popularity with the general public.

In 2019, we have collected almost 220k reports. It represents 100k more than in 2018 and the world coverage is also improved (Figure 5). More than 83% are collected through LastQuake. Although the time distribution depends on seismic activity (Figure 6), the general trend shows that the rate of report collection increase. For instance during the Albanian earthquake sequence in November 2019, we collect 58k felt reports in 7 days… It brokes all records and our growing popularity make us think that it's going to happen more and more often.

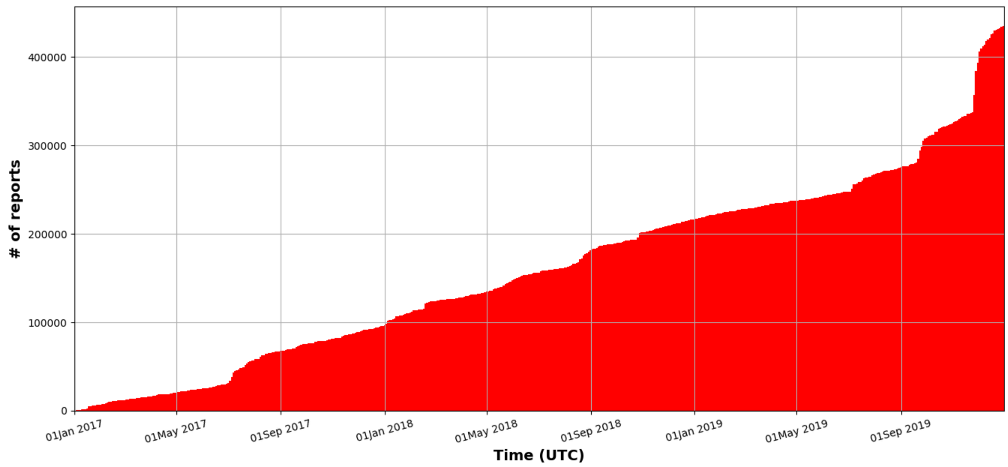


Figure 6: Monthly cumulative time distribution of felt reports collected by the EMSC.

Eyewitenesses often share their feelings with **comments**, especially during seismic sequences. For instance in 2019, we have collected 89k comments, it represents 40% of averall report. **Pictures are a lesser popular way to share experience after an earthquake.** However the number of collected picture increase. In 2018 we have 229 pictures associated to 48 events. These number are outnumbered by the 729 pictures collected in 2019 (associated to 67 seismic events). This is mainly due to the main shock of the Albanian eartquake sequence in November 2019 where 395 pictures have been collected.

3 Traffic monitoring of EMSC dissemination system

On one hand, the dissemination part targets **seismologists and researchers** with dedicated web pages (e.g. "for seismologists only page") accessible trough EMSC websites, the Seismic Portal and its visualization capacities and interactive accesses, webservices. On the other hand, the EMSC targets also **general public and eyewitnesses** with communications on social networks like Twitter, and with the EMSC mobile aplication LastQuake.

3.1 Social networks

The EMSC is present on Twitter, Facebook and Telegram. The main media remains Twitter with 128k followers. For EMSC, Twitter is the media for automatic publication that inform in real time for felt earthquakes. It has to be compared with the 31k Facebook fans and 138 members on Telegram.

Moreover, the number of Twitter followers is constantly increasing. In the beginning of 2020, @Lastquake account reaches 128k followers after an increase of 55% compared to 75.5k followers in August 2018. The monthly evolution of views of EMSC tweets shows same trends where the EMSC popularity on Twitter counts in million (Figure 7).

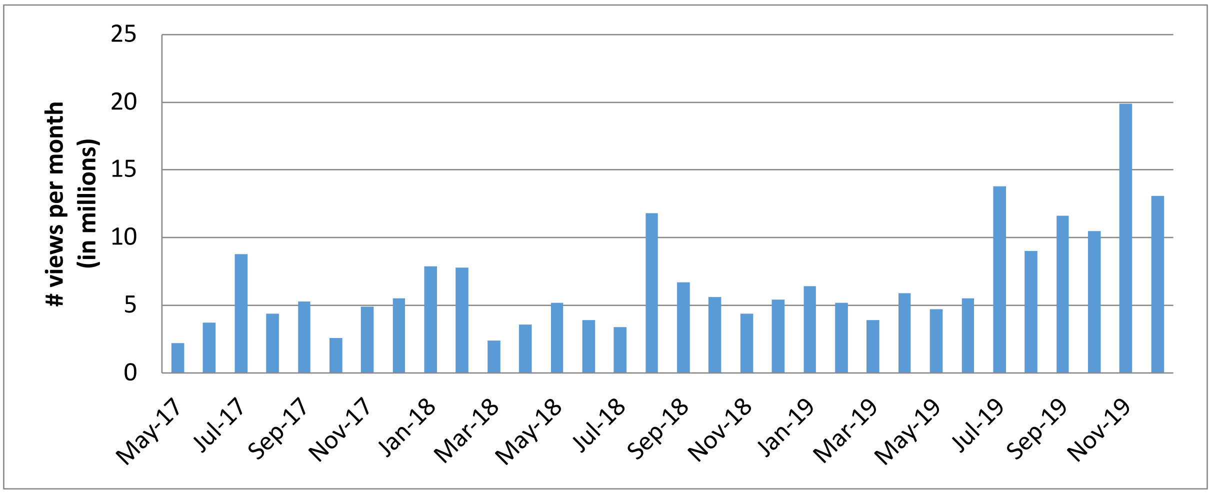


Figure 7: Time distribution of the number of views per month on the @LastQuake Twitter account of EMSC.

3.2 Mobile application: LastQuake

The LastQuake mobile application is an important component of the EMSC communication system. And since September 2018, the number of LastQuake applications in operation has been significantly increased to finally reach 983k in the beginning of 2020 with almost 65% for Android and the rest for iOS (Figure 8).

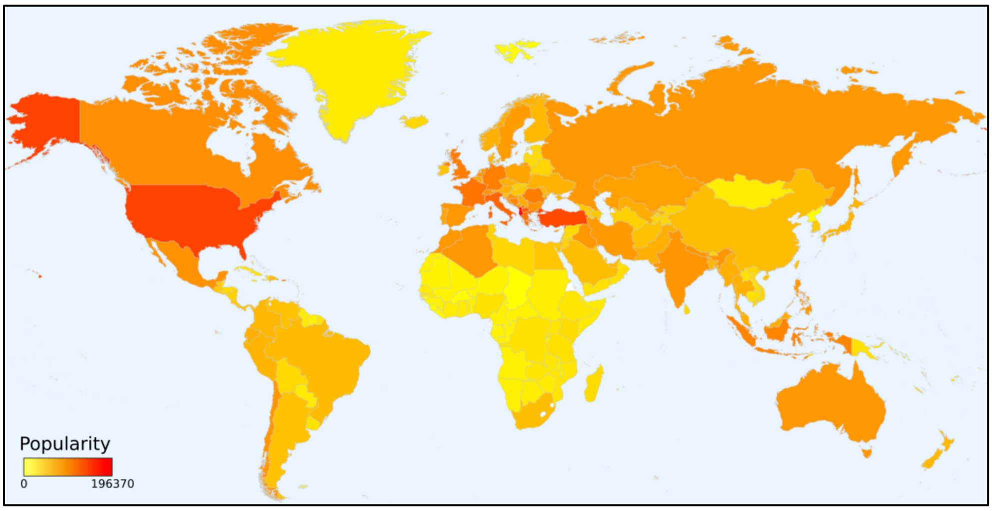


Figure 8: Spatial distribution of LastQuake users in late January 2020. The maximum number of application in activity is reached in Albania.

The time distribution of active users follows the same trend and during the Albania sequence almost 40% of total users were active (Figure 9)! At each felt seismic sequence, we observe that our mobile application becomes more and more popular. Compared to Twitter or websites (see below in section

3.3), number may seem less important… However the mobile application allows a direct communication between the user and EMSC. Since each user is a potential witness of seismic activity, the engagement of user is higher than other media for sharing their experience… as evidenced by the fact that we collect more reports through LastQuake.

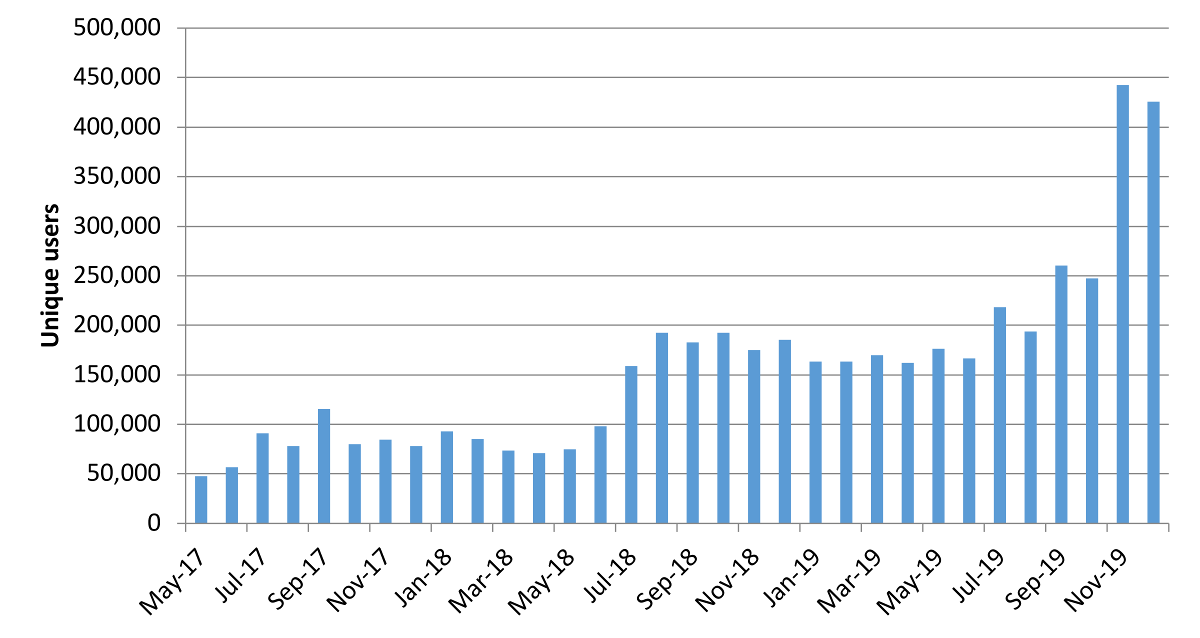


Figure 9: Time distribution of unique users per month that have launched the LastQuake application.

For EMSC, LastQuake is a worldwide success. The application is translated in 16 languages and is used in almost every country (Figure 8). The monitoring of LastQuake launches showed that it has become a system able to detect significant earthquakes in most regions.

3.3 Desktop and mobile Websites

Desktop and mobile EMSC websites (www.emsc-csem.org and http://m.emsc-csem.org) are the traditional place to find our real time data, to search for earthquakes information and to have more information about EMSC. Contrary to Twitter or the mobile application LastQuake, websites are often used by seismologists, in particular the “for seismologists only page”[5].

As shown by traffic curves (Figure 10), they continue to be widely used with more than 800k unique visitors per month for desktop website and almost 200k unique visitors per month for mobile website. All traffic statistics estimations are performed with StatCounter[6]. During the end of 2019, we observe that websites are still major communication media in the EMSC strategy with 3 to 6 times more traffic than usual. This validates our effort to improve our mobile website.

[5] https://www.emsc-csem.org/Earthquake/seismologist.php
[6] www.statcounter.com

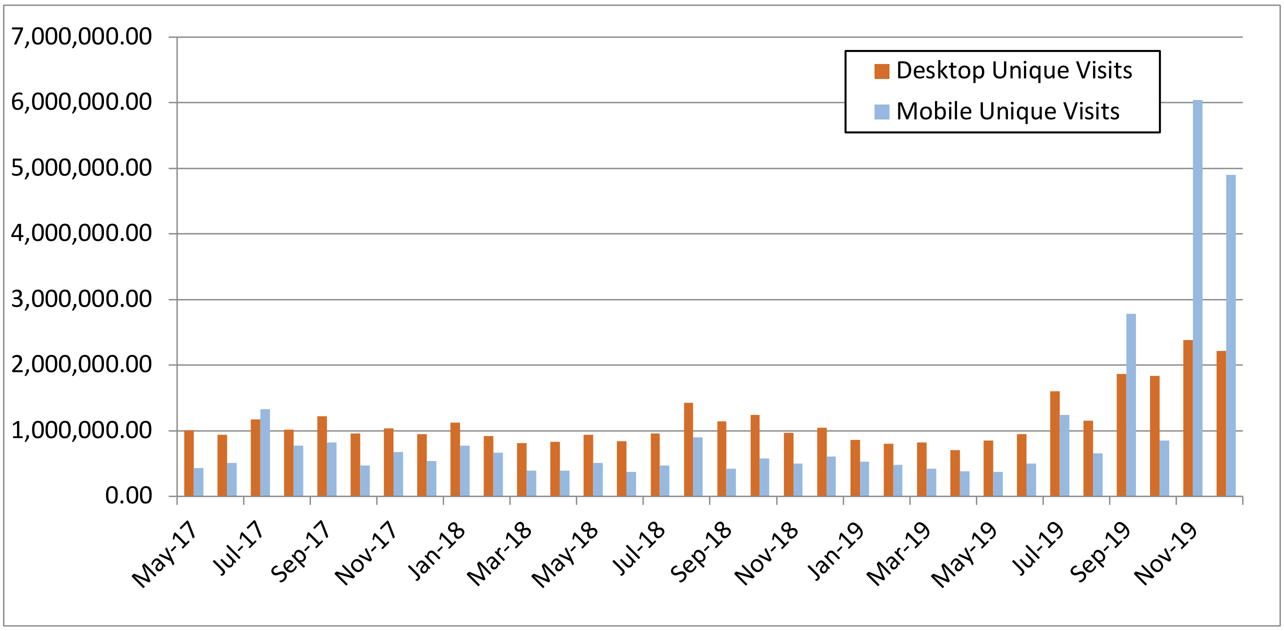


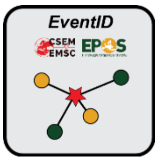
Figure 10: Traffic distribution of monthly unique user on the desktop and the mobile website of the EMSC.

3.4 Seismic Portal

The Seismic Portal is the EMSC portal that gives access to seismic data (www.seismicportal.eu). Users have the choice to use graphical interface or web services for automatic processing.

Most of them have been developed within the European EPOS project and are now operational. To learn how to use them, some use cases are available at https://github.com/EMSC-CSEM/webservices101.

On the Seismic Portal, the EMSC provides 7 services listed below:

	The EventID web service maps dynamically event identifiers to allow the identification of a same event between different seismological institutions.
	Web service conforms to the FDSN-Event standards and providing all the EMSC event data available. Event information can include all origins and all arrivals as desired.
	The service identifies the Flinn-Engdahl region from a geolocalisation entry point.
	Web service that gives access to the moment tensors collected at EMSC

	This service allows downloading all felt reports collected from eyewitness during earthquakes through EMSC websites and LastQuake mobile application.
	Near realtime notification of new and updated earthquake event can be received using the WebSocket protocol. Any WebSocket client can connect to our service to be notified. Javascript, Python example codes are provided to demonstrate the service.
	The web service allows recovering all rupture models from the SRCMOD database of Martin Mai (which is the database of finite-fault rupture models of past earthquakes). These earthquake source models are obtained from inversion or modeling of seismic, geodetic and other geophysical data, and characterize the space-time distribution of kinematic rupture parameters.

The **traffic monitoring of Seismic Portal** services is measured in terms of “hit per month” and “unique IP per month”. We use logs from our front end server to generate traffic measurements. From a global point of view, we count a total of 2 million 188 thousand unique IP addresses in 2019. It represents almost 19k unique users per day and 608k hit per month. To estimate the spatial distribution of Seismic Portal users we use the location of IP addresses thanks to Netacquity services. Although the IP is not always reliable we observe that main users are from Brasil (15%), US (14%), Germany (7%), Chile (7%) and Italy (6%).

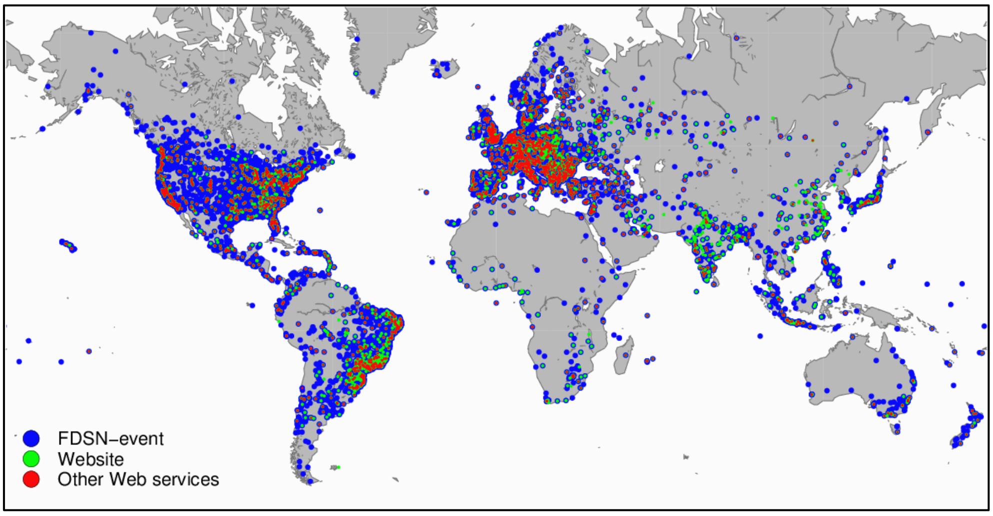


Figure 11: Spatial distribution of Seismic Portal users in 2019.

The **FDSN-event service** is the older service available and it is operational since March 2014. In 2017, the service has gained a large popularity with almost 400k unique users per month (Figure 12). The volume of data transferred via the service is of the order of a hundreds of gigabytes per month. Fdsn-event webservice is also the most popular service of the Seismic Portal with 98% of the overall traffic.

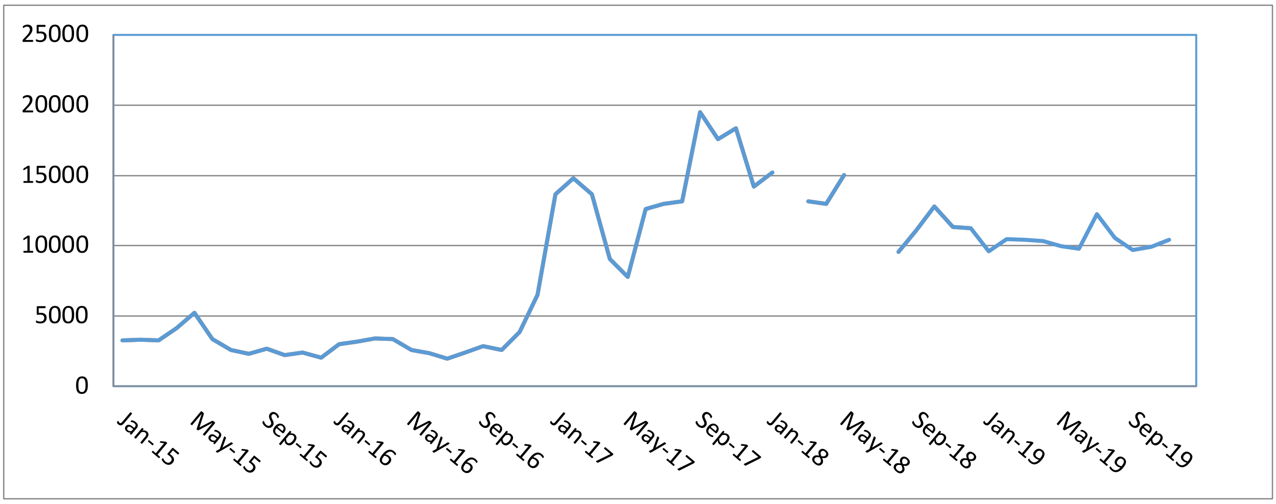


Figure 12: daily unique IP of the FDSN-event webservice available on the SeismicPortal.

Discontinuities of our measurements are due to infrastructure upgrades (beginning of 2018) and a configuration update at the end of May 2018. Moreover the huge traffic generated by the Albanian seismic sequence has overloaded part of our system. We lack of measurement during these 3 periods.

For the other services we have traffic measurements since only August 2018. Of course, they are new and we don't expect traffic numbers to be comparable as for the FDSN-event service (Figure 13). Moreover, these new service provide more specialized data that interested mostly seismologists. In terms of among of data, the felt report service reaches 100Mo in November 2019. The third popular service is the moment tensor service with almost 10Mo of data per month.

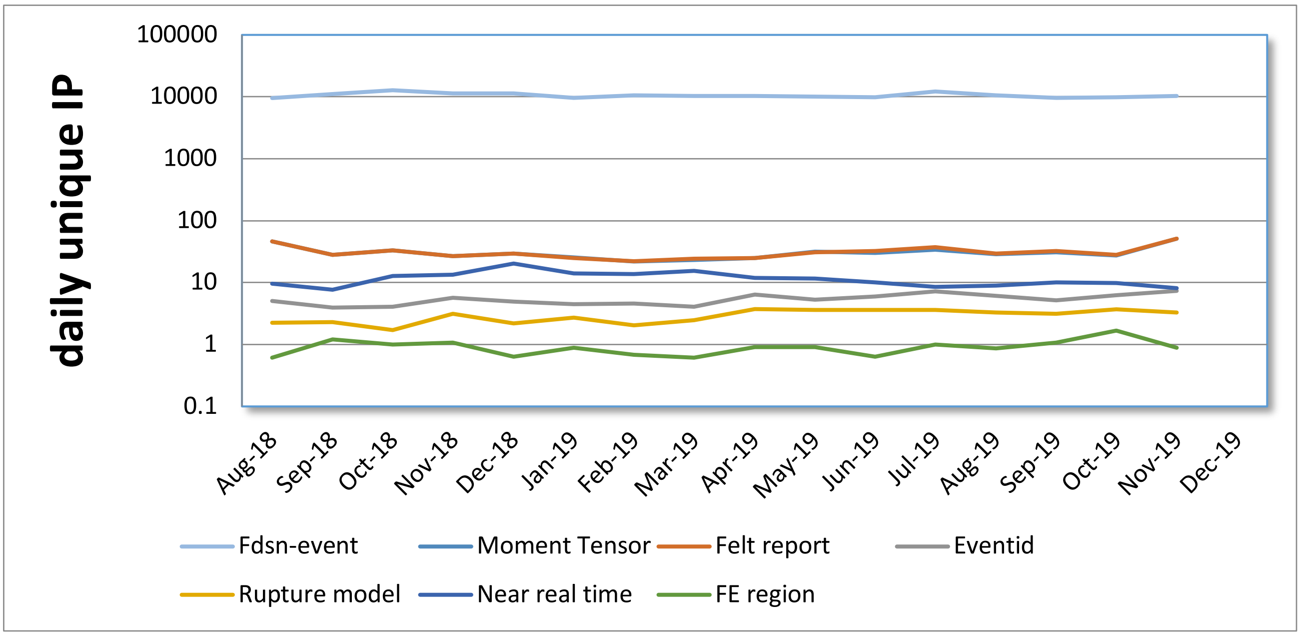


Figure 13: Time distribution of daily unique ip address per Seismic Portal web service.

4 Upgrade of the EMSC core system

All statistics clearly confirm the increasing popularity of our services. And it's a need for us to have a system that can deals with huge traffic peaks like the ones we had during the Albanian seismic sequence. In the earlier phase of the SERA project, we were aware of this necessity and we have dedicated a lot of effort to upgrade the EMSC system. It includes **a hardware upgrade of production servers** thanks to CEA funding and it includes also **a rewrite of its software toolchains**.

Since September 2018, we have upgraded almost all production servers. Most of them are now virtual machines and they are running on recent Dell FX2 servers. More work on the hardware is ongoing with the installation of a backup server and with the use of a new filer.

Concerning the real time system that process seismic information, a lot of effort has been spent to modernize the core system. Beside the code used in production, we have refactored a development version of the system. It is versioned with git[7] and it has been installed on our backup server at the Instituto Geografico Nacional in Madrid. This stage is only a first step and we are now focusing on optimization and reliability of seismic data processing.

[7] https://www.gnu.org/software/git/

5 Conclusion

EMSC gives access to all collected seismic and eyewitness data and takes care of traffic measurements of its services. The traffic monitoring shows the overall popularity of the EMSC among the general public and the seismological community. Compared to the previous period, this popularity is increasing and can reach high traffic peak when earthquake sequence occurs near populated regions like the Albanian sequence in November 2019.

This trend challenges our system and shows key observations.

- Twitter, mobile and desktop website and LastQuake represent the main audience of the EMSC communication strategy and it’s important for EMSC to adapt the message and the information to the general public and we plan to finalize our mobile website and to review our Twitter communication.
- From a scientific point of view, web services are key services to share our data. FDSN-event web service is mature with millions of users. The others services are not as well-known but more and more collaboration are requested to receive felt reports in real time for studying look how to include felt reports in shakemaps and how to constrains rupture models. These web services are important for the EMSC strategy and need to be consolidated.
- Finally, the continuous traffic increase challenges the efficiency of the data processing workflow and we continue our background effort on hardware upgrade and on the modernization and optimization of our seismic tool chain.

SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D20.1 Report on access statistics and service provision of VA1-VA5

(PART 2/ORFEUS-KNMI)

Work package	WP19 VA2: Access to seismic waveforms at ORFEUS/KNMI
Lead	KNMI
Authors	Reinoud Sleeman, Jarek Bienkowski - KNMI
Reviewers	R. Bossu (EMSC), K. Saleh (ETH)
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	29.02.2020
Submission date	30.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D20.1_Report_Virtual_Access_statistics.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

This version of the report is an update of the contribution of VA2 to deliverable D18.1 'Report on access statistics and service provision of VA1-VA5' M16 report and takes into account the review and recommendations of the deliverable D18.2 'Assessment report of VA External Board on VA1-VA5' with respect to VA2.

This report summarizes the provision of the main European services for virtual access to seismic waveform data, related metadata and products for seismology and engineering seismology, through the ORFEUS infrastructure. Through the use of standardized services (e.g. webservices) and GUI's our virtual access services are aligned and compatible with EPOS and EPOS ICS. Moreover, interactive services like StationBook and RRSM have been re-designed with modern development technologies to offer flexibility to add new functionalities (e.g. new types of data or metadata) and tailor it to on-going, changing requirements.

Important updates are:

- In January 2020 the ORFEUS EIDA infrastructure services are fully transitioned to internationally standardized webservices, to became independent of the previous generation of services based on ArcLink.
- Updated statistics on the use of different services are provided.
- ORFEUS EIDA started to deploy the EIDA Federator, developed and deployed by ETH, to offer a one-stop-shop for all EIDA data holdings.
- The ORFEUS Data Center moved to Cloud Computing, providing a platform for storing, accessing and processing data over Internet.
- Further expansion of EIDA with two new nodes - University of Bergen and Institut Cartografic Geologic de Catalunya - and an overall increase in number of stations available from EIDA, from about 8000 to more than 11000.

1 Introduction

The ORFEUS infrastructure is one of the largest infrastructures in the world that provides seismological data and derived products to the scientific research community in strong collaboration with European seismological observatories. The infrastructure is organized as a networked system of observatory infrastructures, waveform data archives and services. A key component is the federated, distributed European Integrated waveform Data Archive (EIDA) that transparently connects a number of large data centers in Europe, including the ORFEUS Data Center. This unique, federated archive serves seismological waveform data from permanent and temporary networks of broad-band sensors and strong motion sensors deployed in Europe and beyond through dedicated services. Currently, EIDA holds beyond 400 TB of data of 107 permanent networks and 190 temporary networks, with a total of more than 11000 seismic stations (status of 25 Feb 2020). Through EPOS-IP, in which ORFEUS is strongly involved, and being compatible with EPOS ICS we foresee that EIDA will extend to serve other data types (OBS, NFO) to a broader user community (e.g. earthquake engineering). At the same time ORFEUS EIDA is a key player in the international Federation of Digital Seismograph Networks (FDSN) and actively contributes to the development of standards and new services. As an example, ORFEUS recognizes the need for - and actively supports developments towards - the standardization of services and formats (e.g. StationXML) in the FDSN and takes. Together with IRIS ORFEUS takes responsibility in the development of the FDSN webservice 'data_availability', as well as in the creation of future proof documentation on StationXML.

Services that are being offered to the (seismological) research community to provide (virtual) access to raw waveform data and related metadata are: a) ORFEUS website, b) EIDA data portal, c) EIDA webservices, d) RRSM (Rapid Raw Strong Motion database) and e) StationBook.

Specific objectives of this work package as described in this report are:

- Coordination with NA2 activities on extending EIDA to support other types of data and to serve a broader geoscience community and the engineering and hazard communities (e.g. Implementation of data model extension; Near Fault Observatories or structural monitoring arrays).
- Services offered and developed by this infrastructure for: flexible and transparent access to raw waveform data in EIDA; related metadata (station and data quality); derived data and products; refined data discovery across EIDA; documentation and outreach.
- Access through the EIDA GUI and standardized webservices.
- Access to EIDA StationBook and the RRSM.

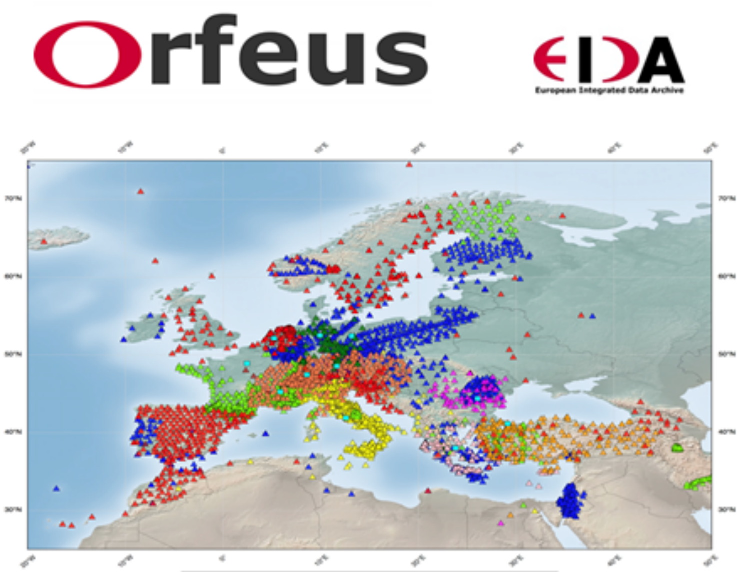


Figure 1: The ORFEUS / EIDA infrastructure provides access to raw seismological waveform data and related metadata from more than 11000 sensors throughout Europe and beyond. The federated data archive EIDA connects 10 large data archives (holdings are geographically identified by colour).

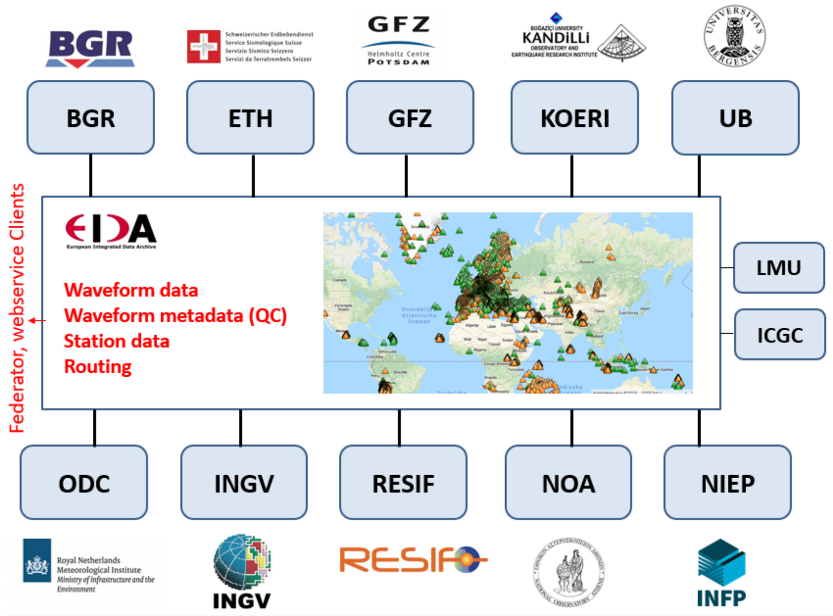


Figure 2: Standardized services that are deployed across EIDA to provide uniform and transparent (virtual) access to all data holdings.

The number of data archives in EIDA was expanded with the entry of University of Bergen (UB). UB is offering all required services and tools that are required by EIDA and serves data from the Norwegian National seismic network as well as from the NORSAR archive. Also, the integration of ICGC (Institut Cartografic Geologic de Catalunya) has started in close collaboration with ODC. ICGC will bring another 50+ seismic stations into EIDA. The (political and technical) work towards the close collaboration with UB and ICGC and the integration in EIDA has been done in close cooperation with the work and activities in SERA NA2.

2 ORFEUS/EIDA services

The services provided can be divided in the categories webservices and interactive services, which are listed here and further described in chapters 2.1 and 2.2.

1. Webservices
   1.1. fdsnws-dataselect
   1.2. fdsnws-station
   1.3. eidaws-routing
   1.4. eidaws-wfcatalog

2. Interactive services
   2.1. ORFEUS website
   2.2. EIDA GUI
   2.3. RRSM
   2.4. StationBook

3. In addition to the above services a number of clients have been developed at ORFEUS Data Center (ODC). These will be described in section 2.3.

4. Federator
   The EIDA federator provides a single, unified access point to the waveform archives and the station and quality control information from the entire EIDA data holdings, i.e. from all the datacenters in EIDA. Access is through standard FDSN and EIDA web services as described in section 1.

2.1 Webservices

ORFEUS EIDA implements the following webservices to provide standardized and open access to seismological (waveform) data. The specifications and the usage of parameters of each service can be found at the appropriate page (through the link):

- fdsnws-dataselect - FDSN standardized webservice for mini-SEED waveform data.
- fdsnws-station - FDSN standardized webservice for station metadata.
- eidaws-routing - EIDA standardized webservice for routing between EIDA services.
- eida-wfcatalog - EIDA standardized webservice for waveform metadata.
- eida-federator - EIDA Federator webservice for transparent data access across EIDA.

ORFEUS EIDA consists of multiple data centers with unique data holdings and webservices. Data exposed at one data center may not be available at another, therefore the appropriate node should be selected in your request. Please consult the EIDA networks page to discover the appropriate node(s) for data requests and citation.

The new EIDA service, called Federator (http://www.orfeus-eu.org/data/eida/nodes/FEDERATOR/), designed and deployed at ETH, supports fdsnws-station, fdsnws-dataselect, and eidaws-wfcatalog requests across all EIDA nodes to enable users to collect data without *a-priori* knowledge of where data is hosted.

Within EIDA two other services are being developed and tested in order to optimize harvesting of data in a complex, federated system like EIDA and to enable users to easily collect open and restricted data:

- the EIDA mediator webservice will be designed for advanced selection of data across EIDA based on user criteria (e.g. quality parameters).

- the EIDA authentication webservice is a central authentication system that provides tokens for all EIDA services across EIDA. The authentication System connects to a B2ACCESS service (provided by the EUDAT Collaborative Data Infrastructure). This webservice is being tested.

	FDSNWS-Dataselect	FDSNWS-Station	EIDAWS-WFCatalog
FEDERATOR	Online 1.1.0	Online 1.1.0	Online 1.0.0

Table showing current status, service versions, and response times for the FDSN and EIDA Webservices. **Visit the individual node pages to discover the webservice addresses for each node.**

EIDA Node	FDSNWS-Dataselect	FDSNWS-Station	EIDAWS-Routing	EIDAWS-WFCatalog
ODC	Online 1.2.0	Online 1.2.0	Online 1.2.0-b4	Online 1.0.0
GFZ	Online 1.2.0	Online 1.2.0	Online 1.2.1-b1	Online 1.0.0
RESIF	Online 1.1.0	Online 1.1.0	In development	Online 1.0.0
INGV	Online 1.2.0	Online 1.1.41.1	Online 1.0.4	Online 1.0.0
ETHZ	Online 1.2.0	Online 1.2.0	Online 1.2.0-b1	Online 1.0.0
BGR	Online 1.2.0	Online 1.2.0	Online 1.1.0	Online 1.0.0
NIEP	Online 1.2.0	Online 1.2.0	In development	Online 1.0.0
KOERI	Online 1.2.0	Online 1.2.0	Online 1.1.1	Online 1.0.0
LMU	Online 1.2.0	Online 1.2.0	Online 1.0.3	Online 1.0.0
NOA	Online 1.2.0	Online 1.2.0	Online 1.2.0-b4	Online 1.0.0
UIB	Online 1.1.0	Online 1.1.2	In development	Online 1.0.0

Figure 3: Status of the webservices across EIDA. Each EIDA node runs the same services (fdsnws-dataselect, fdsnws-station and eidaws-wfcatalog) and optional the routing service. When consulting this webpage (https://www.orfeus-eu.org/data/eida/webservices/) each service is queried across EIDA in order to reflect the current status. Also the status of the new service 'Federator'is shown on top.

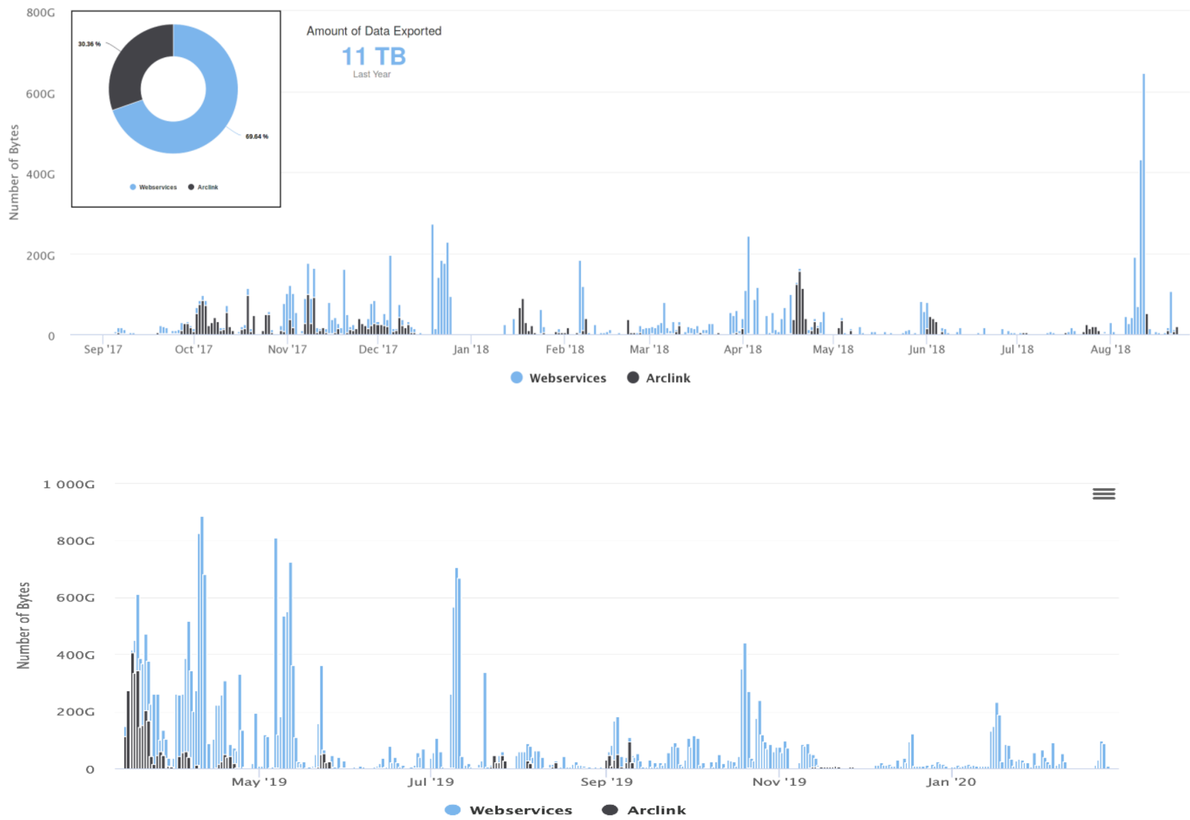


Figure 4: Amount of data (GBytes) exported by ODC between Sep 2017 and Augustus 2018 (top; total 11 TB) and between April 2019 and February (bottom; total 28 TB), through webservices and ArcLink (which is phased out).

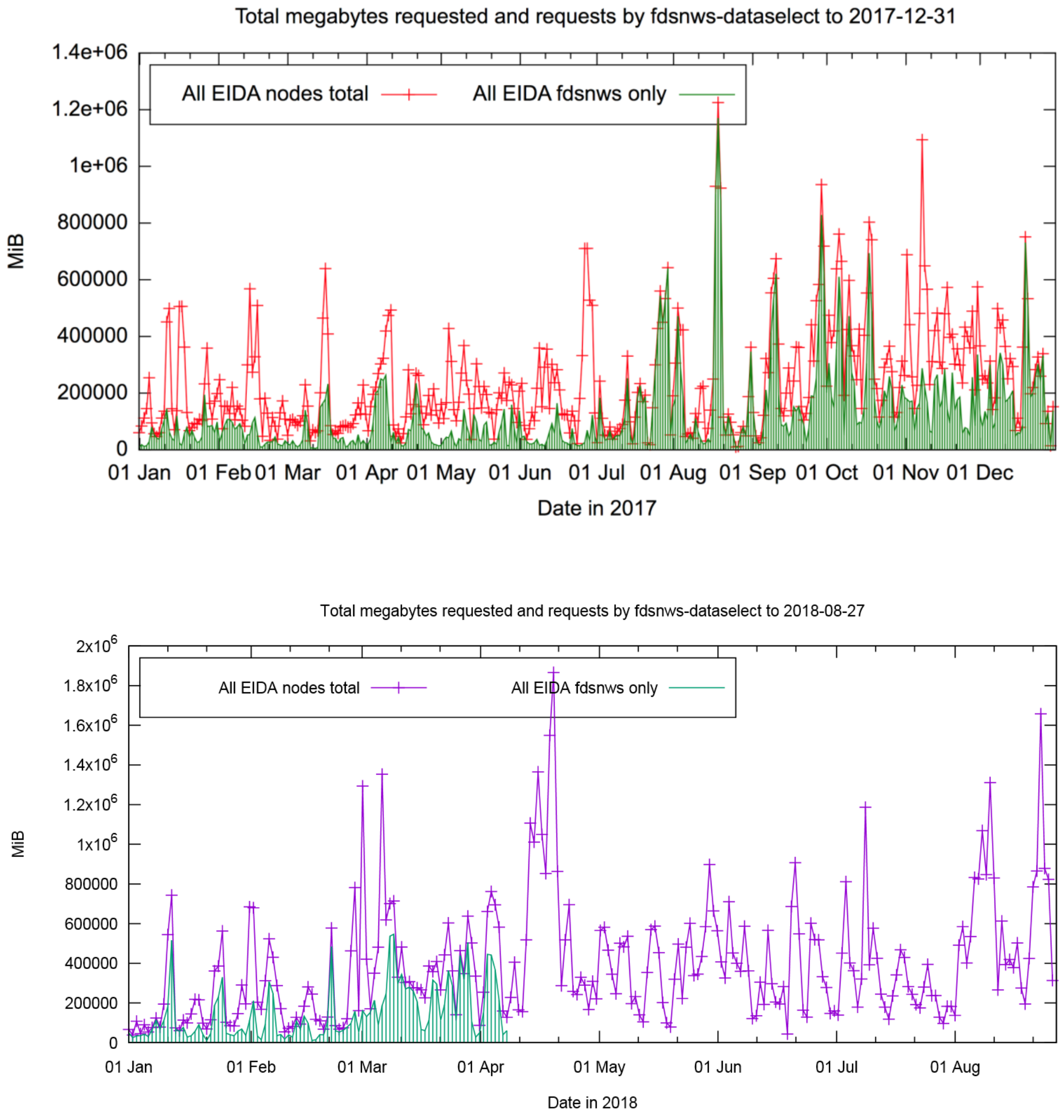


Figure 5: Amount of data (Bytes) exported by EIDA throughout 2017 and 2018.

Since the beginning of 2020 the service ArcLink has been phased out across EIDA. This has been communicated widely across the research community and the use of webservices has been advertised on the ORFEUS website (http://www.orfeus-eu.org/data/eida/). See the Advanced Workflow Examples page or the ORFEUS EIDA Webservices Notebook for a more extensive explanation on how to use the webservices instead.

2.2 Interactive services

Interactive services that are provided by ORFEUS EIDA to serve waveform data and information are :

- ORFEUS website - the landing pages for all information concerning ORFEUS, EIDA and services.
- EIDA web interface - the GUI to interactively search for and download data from EIDA.
- StationBook - the GUI to access all (available) information on seismic stations across EIDA.

- RRSM web interface - the GUI to search for and collect strong motion products in near real time.

2.2.1 ORFEUS website

The ORFEUS website is the entry point for the variety of services and information to collect seismological waveform data hosted by European data archives.

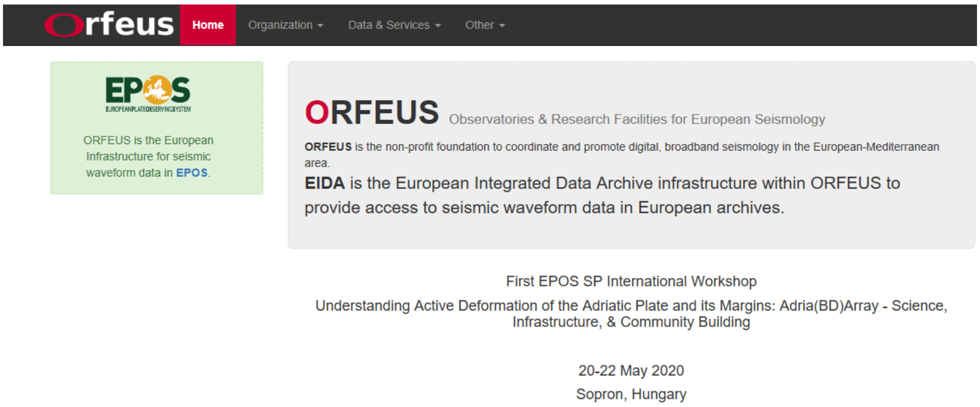
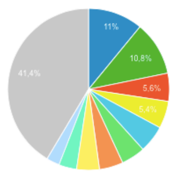


Figure 6: The ORFEUS website 'landing page'.

Land	Gebruikers	Gebruikers
	15.230 % van totaal: 100,00% (15.230)	15.230 % van totaal: 100,00% (15.230)
1. United States	1.682	10,97%
2. Italy	1.660	10,82%
3. Czechia	861	5,61%
4. France	835	5,45%
5. Greece	833	5,43%
6. China	734	4,79%
7. Netherlands	731	4,77%
8. Germany	697	4,55%
9. Turkey	544	3,55%
10. Russia	404	2,63%

Bijdrage aan totaal: Gebruikers



Land	Gebruikers	Gebruikers
	35.245 % van totaal: 100,00% (35.245)	35.245 % van totaal: 100,00% (35.245)
1. United States	5.515	15,24%
2. Italy	3.364	9,30%
3. China	2.113	5,84%
4. Turkey	2.093	5,78%
5. Greece	1.808	5,00%
6. France	1.624	4,49%
7. Germany	1.453	4,02%
8. Netherlands	1.315	3,63%
9. Japan	1.133	3,13%
10. United Kingdom	924	2,55%

Bijdrage aan totaal: Gebruikers

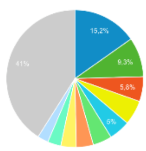


Figure 7: Numbers and percentages of different users invoking the ORFEUS website, in the first reporting period (top) and from 01-01-2018 – 25-02-2020 (bottom). Notice the impact of ORFEUS outside of Europe (e.g. United States, China, and more recently Japan).

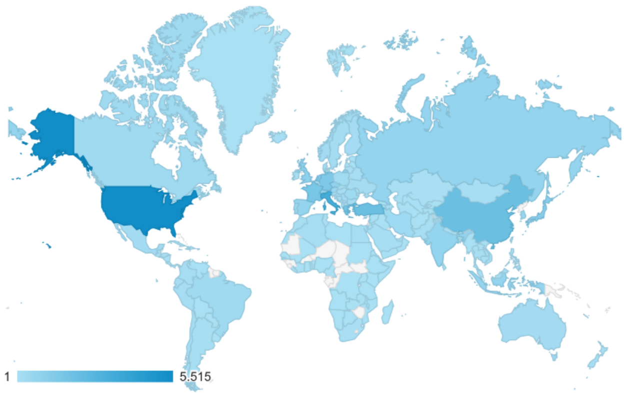


Figure 8: Geographical distribution of users across countries (01-01-2018 – 25-02-2020). Notice the global impact of the services provided by ORFEUS.

2.2.2 EIDA webinterface (GUI)

The EIDA webinterface, originally developed by GFZ, is installed at a number of EIDA nodes and demonstrates the transparency for the community to search for and download data from the federated archive EIDA.

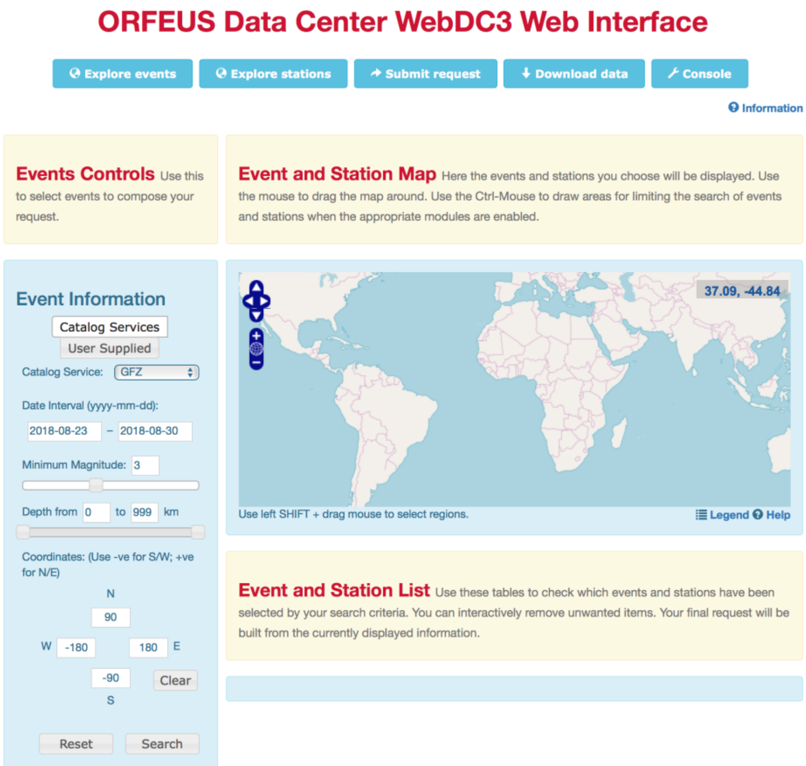


Figure 9: EIDA webinterface.

A new design of the EIDA interface was developed and put in place to provide an EIDA access point allowing users to browse stations, events and request waveform data using extensive filtering mechanisms. By design one of the aims of the new EIDA interface is to make it modular and open for extensions (e.g. new data types and metadata models). With modern development methodologies favouring micro-services, loosely coupled modules and separation of concerns it will be easy to add new functionalities and tailor it to rapidly changing requirements.

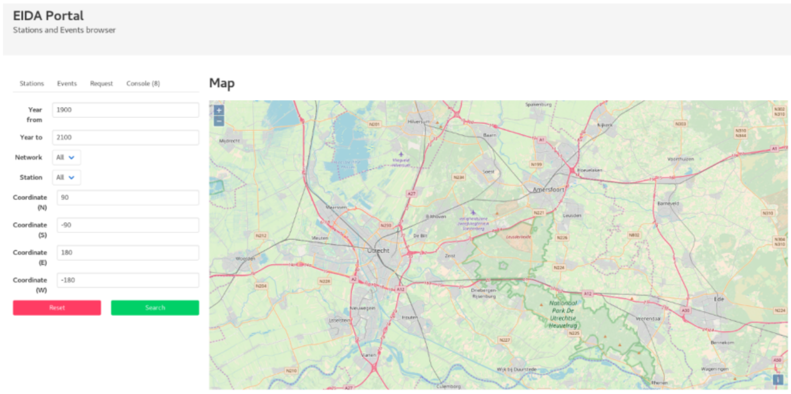


Figure 10. New EIDA interface home page prototype (still in development).

2.2.3 Station Book

Station Book is a web application built on top of FDSN and EIDA web services with its own backend logic and database. It is intended to be the complete, interactive catalogue of EIDA stations containing extended station metadata not covered by the FDSN specification (operator notes, descriptions, photos, comments). Users can register themselves but need to be given write access to a network to be able to start editing its stations. Both backend logic and underlying structure of the StationBook has been redesigned and upgraded significantly (e.g. Python3) this enables flexibility to incorporate new types of extended station metadata (e.g. new database tables). Data ingestion is in two ways:

- common network and station metadata: collected automatically from EIDA.
- station and site characteristics: added/edited by network operators.

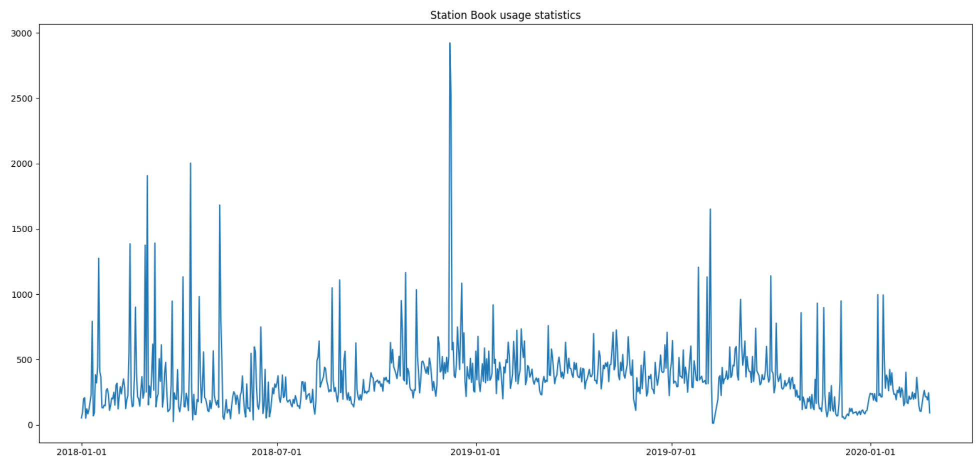


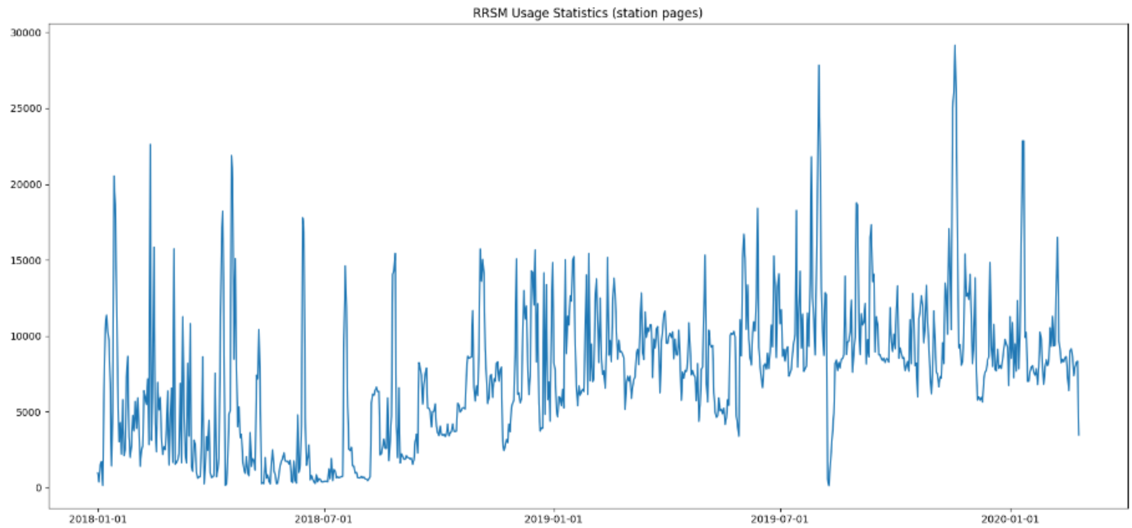
Figure 11. Daily number of requests made to the StationBook (01 January 2018 - 25 February 2020).

As the new StationBook was significantly renewed we provide below the important (technical) changes in the StationBook and a detailed description of the GUI in Appendix 4.1

- Backend written in Django 2.0 (Python 3)
- Frontend in Bootstrap 4 + OpenLayers
- Using Routing WS + FDSN WS from all EIDA nodes for data sync (basic station data)
- Has its own cache table for basic station data (Network, Code, Name, Latitude, Longitude, Elevation, Status, Start, End, Created)
- Channels and instrumentation data obtained in real time from appropriate node via FDSN WS
- Additional data (Owner, Morphology, Housing, Borehole etc.) stored in dedicated tables
- Lots of possibilities for extensions (additional data and functionalities like bulk data upload, comments, etc.)

2.2.4 RRSM

The RRSM portal allows users to query earthquake information, peak ground motion parameters, response spectral amplitudes and to select and download earthquake waveforms within minutes after an earthquake with magnitude ≥ 3.5 occurring in the European-Mediterranean region [1]. Earthquake information is provided by the EMSC and all on-scale seismic waveform data available from ORFEUS EIDA is considered for fully automated processing. Real-time RRSM processing started in June 2014. Offline reprocessing was carried out for all M ≥ 4.5 events that occurred since January 2005, and all M ≥ 3.5 events since January 2012.



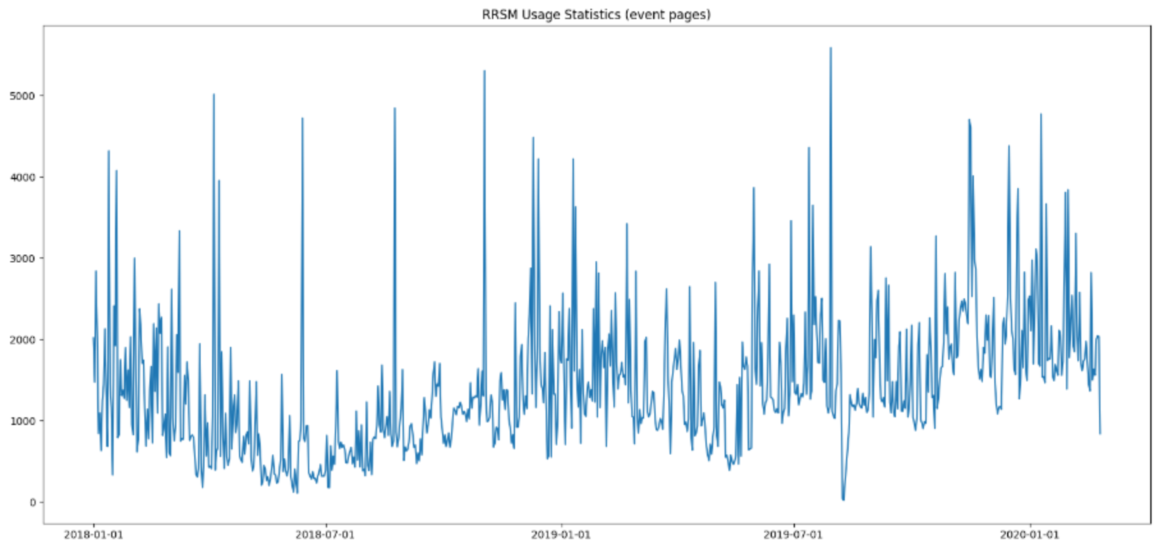


Figure 12: Daily number of requests made to the RRSM (01-01-2018 until 25-02-2020). On top the number of station specific requests made for detailed strong motion values. The bottom graph shows the number of requests for generic event information.

The RRSM GUI is a visual representation of information available in the RRSM database and through RRSM web service. Just as with the StationBook the RRSM backend logic and underlying structure has been redesigned and upgraded significantly within SERA. We provide below the important (technical) changes in the RRSM as well as a detailed description of the GUI in the Appendix 4.2:

- Backend written in Django 2.0 (Python 3)
- Frontend in Bootstrap 4 + OpenLayers for maps
- Using ODCWS RRSM web service

2.3 Clients

ORFEUS Data Center developed a number of specific clients (https://www.orfeus-eu.org/data/odc/) to display features like data latency, event waveforms and data quality parameters.

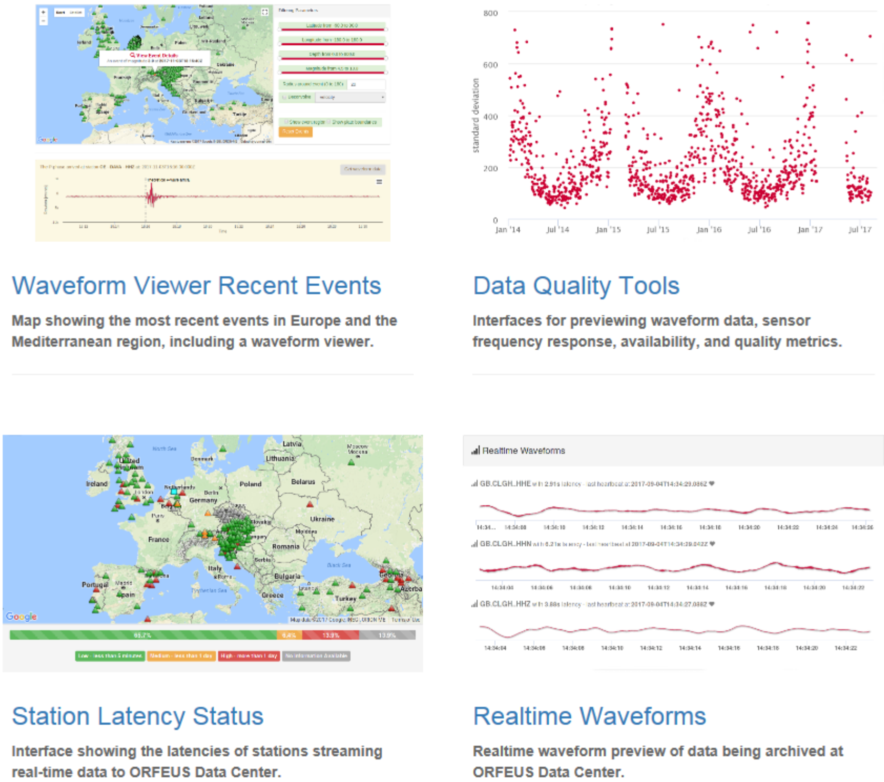


Figure 13: Example clients at ODC displaying a) earthquake waveforms, b) data quality parameters, c) data stream latencies and near real-time waveforms. Also clients to display data availability, instrument responses, data export statistics and statistical values of the data samples are available.

Besides visual interfaces to waveforms, the EIDAWS-WFCatalog [2] provides an API that exposes a waveform metadata catalogue for the seismic archive at an EIDA node. The WFCatalog Webservice provides detailed information on the waveform data like quality parameters (derived from data record headers, e.g. timing quality and header flags) and statistical values (derived from the sample values, e.g. rms). The WFCatalog can serve as an index for data discovery (e.g. Mediator) as it has support for range filtering on all available metrics. The quality parameters are continuously calculated and stored in the WFCatalog database, enabling fast and efficient querying of these parameters (no on-the-fly calculations on the waveforms are needed). This enables for example fast computation of PDF's for seismic data over long time frames (e.g. 1 year).

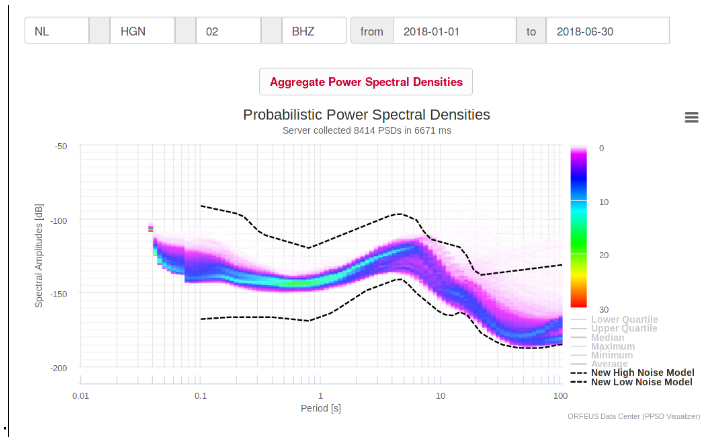


Figure 14: Example of the Probability Density Function calculated from 6 months of continuous seismic data, using the pre-calculated Power Spectral Density values within the WFCatalog database.

2.4 Federator

For ORFEUS EIDA ETH developed the EIDA Federator to offer:

- a one-stop-shop for all EIDA data holdings. The Federator returns all data using a single request, independently of which datacenter is curating it, and whether it is distributed over multiple data centres.
- direct support for virtual networks with unrestricted data access.
- the handling of large data requests by allowing high maximum limits for requests received, not limited by the ones configured at the endpoints. The incoming requests will be split in smaller pieces. Requests to the Federator also benefit from traffic shaping and load balancing across nodes.

The Federator is currently implemented within ORFEUS EIDA to provide the user community with a tool to make a simple and transparent use of near-real time data access or mass downloading of data feasible, especially for current and planned large-N experiments. Currently the service is monitored on performance.

References

[1] Introducing the European Rapid Raw Strong-Motion Database. C. Cauzzi et. al., 2016, Seismol. Res. Lett. 87, 4, doi: 10.1785/0220150271

[2] Trani, Luca & Koymans, Mathijs & Atkinson, Malcolm & Sleeman, Reinoud & Filgueira, Rosa. (2017). WFCatalog: A catalogue for seismological waveform data. Computers & Geosciences. 106. 10.1016/j.cageo.2017.06.008.

3 Appendices

3.1.1 Station Book

The StationBook has 3 types of users:

- Guest users (non-registered users with read only access)
- Registered users (read only access directly after registration)
- Admins (users which can give registered users write access to network/station metadata)

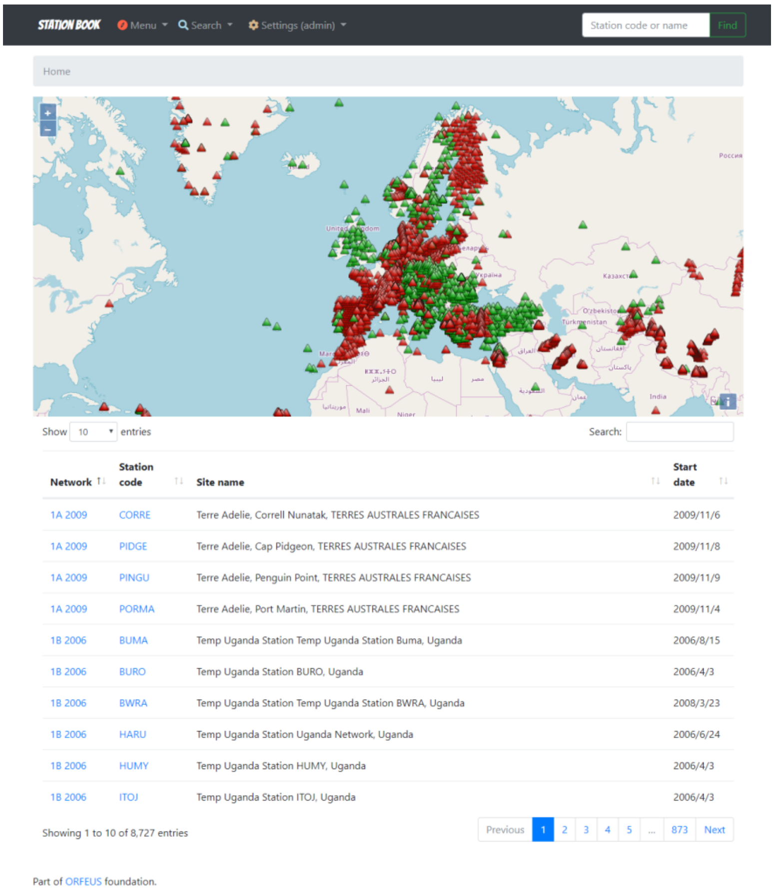


Figure A1: Station Book home page

The navigation bar located at the top of the screen contains a menu which allows users to quickly navigate to one of following pages:

Via "Menu":

- View all nodes (list of EIDA nodes hosting FDSN web services used by the Station Book)
- View all networks (available networks)
- Recent changes (list of recent changes of the stations metadata)
- Quick links
- About Station Book

Via "Search":

- Search (extended station search form)

Via "Settings":

- My account (editable account details e.g. name, e-mail, telephone)
- Change password
- Log out
- Admin panel (only when logged in with administrator rights)

On the right side of the navigation bar a "quick search" textbox can be found which can be used to quickly navigate to station by typing its code or site name.

The networks page contains a list of all networks archived at EIDA. Networks can be quickly filtered using the search box located above the table. The network details page contains network details and lists all stations belonging to this network. Stations can be quickly filtered using search box located above the table.

The station details page presents station information from the FDSN web service and extended information stored within the Station Book.

FDSN information (read only for all users):

- FDSN-Station Data
- FDSN-Station Instrumentation Data

Extended information stored within the Station Book (writeable for users with write access to the network from which given station originates):

- Basic Data
- Owner Data
- Morphology Data
- Housing Data
- Borehole Data
- Photos

Extended information can be edited using "Edit data" dropdown menu located below the map.

Photos can be added via station gallery page which can be accessed using "Media" dropdown located below the map.

The station details page provides also access to two modal windows which can be opened using buttons located on the left side of the station search box in the navbar:

- Information window with explanation of morphology classes, ground types and station sensor types
- Station change log showing recent changes of the station extended information

The station gallery page presents photos uploaded by the station operators. Additional photos can be added using "Upload photos" button located on the right side of the page (available only for users with write access to the network from which given station originates).

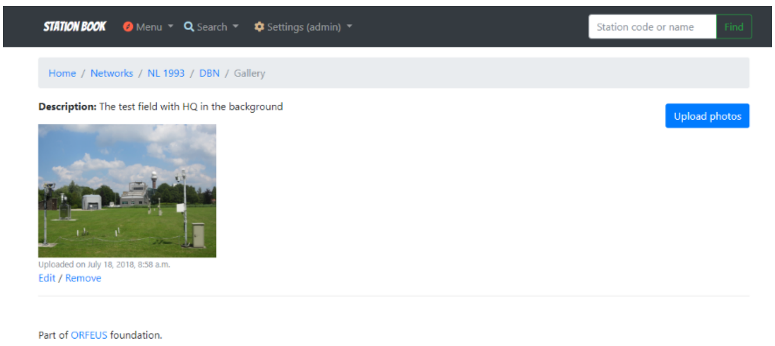


Figure A2: Station Book station gallery page

3.1.2 RRSM

The navigation bar located at the top of the screen contains a menu which allows users to quickly navigate to one of following pages:

Via “Recent events” users can filter events which occurred in various time frames. Menu entries allow users to show events from last 24 hours up to last 10 years.

Via “Search”:

- Search events
- Search peak motions
- Search combined
- Custom search

By default, the home page renders events which occurred in the last month. Clicking on event Flinn-Engdahl region in the “Events” table will focus the map on given event. Clicking on event origin time in the “Events” table will show the event details page.

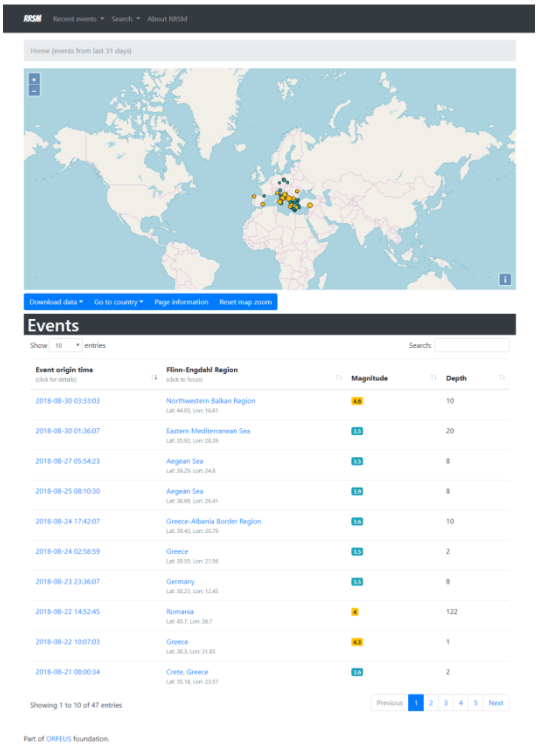


Figure A3: RRSM home page

The search events page allows users to search for events based on event characteristics.

The search peak motions page allows users to search for events based on PGA and PGV characteristics.

The search combined page allows users to search for events based on magnitude, station location and PGA/PGV characteristics.

Custom search page allows users to search for events based on custom characteristics which can be enabled and disabled using the button list located above the form.

Event details page presents the event information obtained via RRSM web service:

- Location of the event
- Locations of stations which have been triggered by the event
- PGA vs epicentral distance graph
- PGV vs epicentral distance graph
- Earthquake information
- List of stations with maximum recorded PGA and PGV values, epicentral distance and elevation

There is a button group below the map which allows user to preview RRSM web service response used to generate the page, download ShakeMap XML and Processed Waveforms, navigate to WebDC3 interface and show page information modal window.

Clicking on the network and station code in the "List of stations" table will navigate to event station stream page.

Event station stream page presents the event station streams information obtained via RRSM web service:

- Location of the event
- Location of given station
- Pseudo-Spectral Acceleration graph
- Displacement Response Spectra graph
- Station information
- List of streams

There is a button group below the map which allows user to preview RRSM web service response used to generate the page, navigate to WebDC3 interface and show page information modal window.

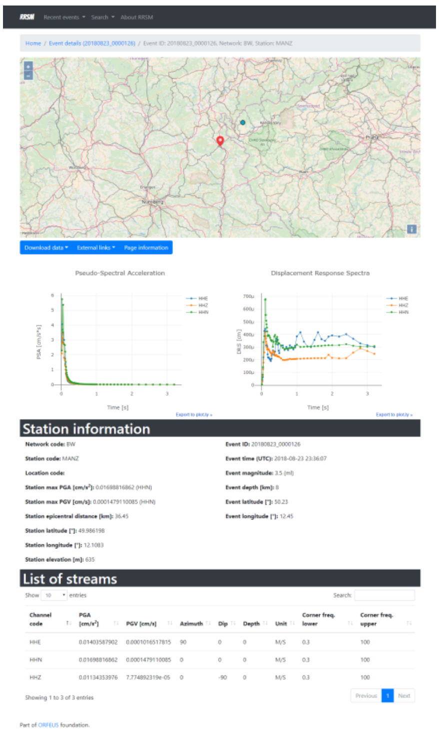


Figure A4: RRSM Event station streams page

SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D20.1 Report on access statistics and service provision of VA1-VA5

(PART 3/INGV)

Work package	WP 20 Access to data and services for engineering seismology (VA3)
Lead	INGV
Authors	L. Luzi, G. Lanzano, M. Locati, A. Rovida, R. Basili, R. Vallone, INGV
Reviewers	Hong Kie Thio, AECOM, R. Bossu (EMSC), K. Saleh (ETH)
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	29.02.2020
Submission date	30.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D20.1_Report_Virtual_Access_statistics.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

The access to data and services for engineering seismology include: i) European Strong Motion Database (ESM); ii) European Archive of Historical Earthquake Data (AHEAD); and iii) European Database of Seismogenic Faults (EDSF). This work package will strengthen and coordinate its currently separated, and intrinsically diverse, services to provide optimized access to data and tools for the seismological and the engineering seismology communities.

This deliverable describes the service provision and the access to the three services, in terms of number and type of users, number of visited pages and data download.

1 Chapter 1: Description of the services offered by the VA3

1.1 Common access to services for engineering seismology

In the former report, we described the common portal that has been created to access data and services for engineering seismology, represented by the three databases described in paragraph 1.1.1 to 1.1.3 (http://sera-va3.rm.ingv.it/, see Fig. 1).

During the last year of activity, there was an increased use of web services for accessing data, as we will demonstrate later on in this report. In order to expand the user base even further, the group decided to investigate the potential of an additional method to access the data that is aimed at users that might find it difficult to interact with web services. The basic idea was to implement a direct access to data for GIS users within their environment in a convenient and user-friendly way. Up to now, GIS users were able to easily include data provided via OGC standards only because they are well supported by GIS software out-of-the-box. Instead, GIS users found quite difficult to include data coming from web services based on other standards, such as those promoted by the International Federation of Digital Seismograph Networks (FDSN). To achieve the goal, at the beginning of 2019 we started drafting a plugin for QGIS, a widely adopted open source GIS software. By design, the plugin takes advantage of the already available SERA-VA3 web services, but instead to tie it up to a specific service, we planned to implement the support for a web service standard. This approach will allow the plugin to be customized by the final user and will allow to include any web service compliant to that particular standard. So far, SERA-VA3 web services deliver data using the following web services standards:

- FDSN-event, providing access to earthquake parameters encoded in QuakeML 1.2;
- FDSN-station, providing information about seismological stations encoded in StationML;
- AHEAD macroseismic service, providing macroseismic data encoded in QuakeML 2.0;
- Open Geospatial Consortium Web Map Service (OGC WMS);
- Open Geospatial Consortium Web Feature Service (OGC WFS);

In November, we started a collaboration with a private company for the development of the plugin. At the time of writing this report, February 2020, the plugin functionalities are properly outlined, but only a small portion of them are implemented, therefore the plugin should be considered in a pre-alpha state (Fig. 2). The plan is to release a beta version by mid-March 2020, and the source code in a public development environment (i.e. GitLab or GitHub) so that anyone willing to contribute to the plugin will be enabled. So far, we have found a positive reaction from a small group of people we interviewed, and we hope that in the last month of the project activity, we will be able to provide a feedback from a wider user base.

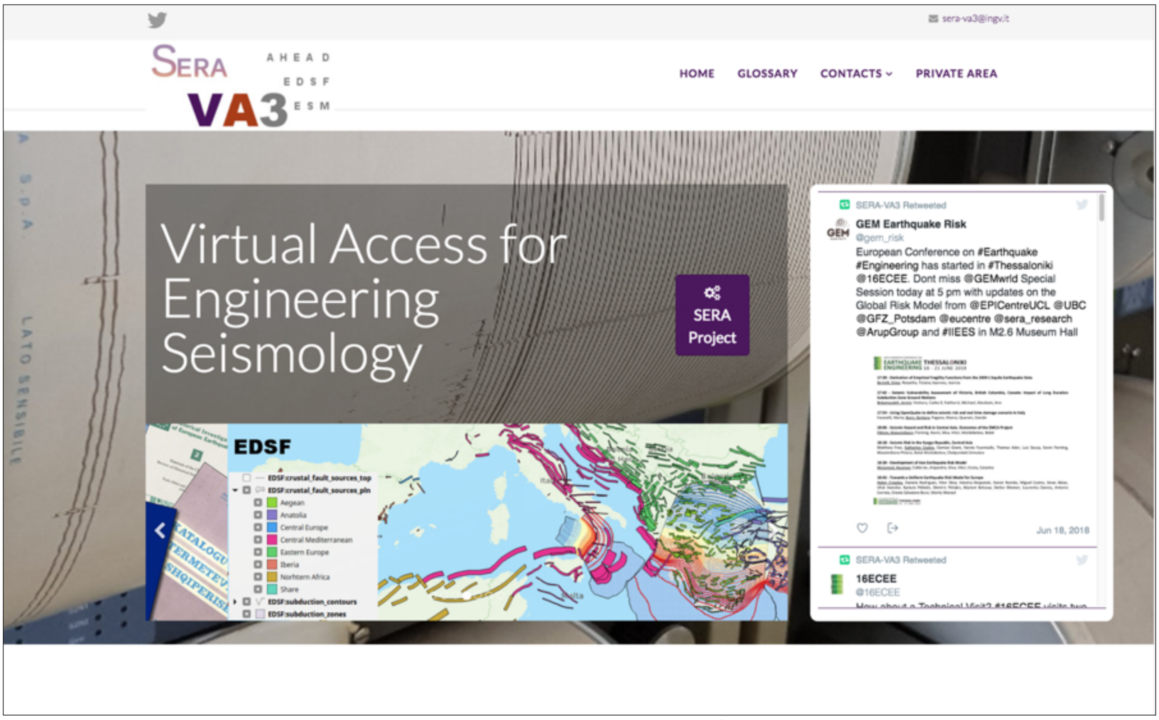


Figure 1: SERA-VA3 portal

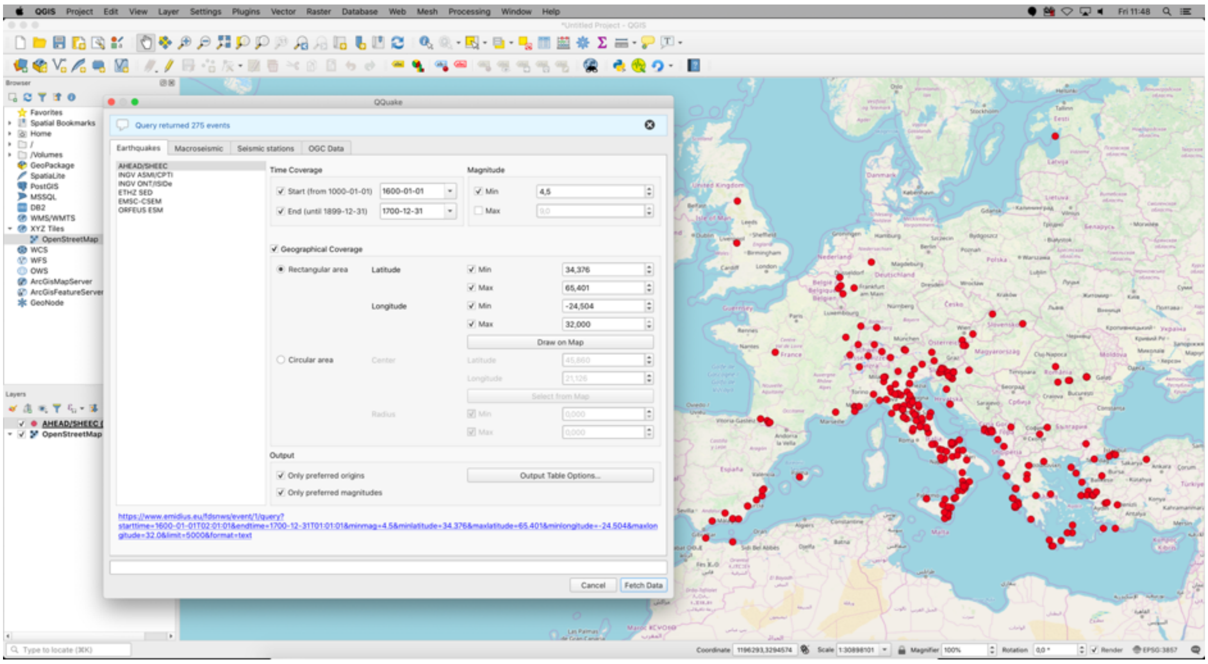


Figure 2: Pre-alpha version of the plugin for loading SERA-VA3 data in the QGIS software.

1.2 European Strong Motion database (ESM)

ESM is a centralised collector of European Strong motion data, with magnitude threshold of seismic events equal to 4. It archives the waveforms recorded since 1969 by about 50 European seismic networks and provides end-users with quality-checked and manually processed waveforms. The database is updated daily with new waveforms and metadata and the number of available waveforms is about 68000 at the end of January 2020. The service is distributed and regulated under the umbrella of ORFEUS (Observatories & Research Facilities for European Seismology, www.orfeus-eu.org/) and is contributing to the "seismology" pillar of the European Plate Observing System (EPOS, https://www.epos-ip.org/).

The access to data is provided through a web interface (http://esm.mi.ingv.it) and several web services that have been published in 2018 and 2019:

1) https://esm.mi.ingv.it//esmws/eventdata/1/ provides machine-friendly access to event based waveforms; this service is accessible to registered users.

2) https://esm.mi.ingv.it//esmws/shakemap/1/ provides machine-friendly access to the USGS ShakeMap input for calculating the shakemap of a single event; the input file contains the peak values and the acceleration response spectra ordinates at three periods (0.1s, 1s and 3s), relative to the records of one seismic event.

3) https://tex.mi.ingv.it/esmws/rexel-target-spectrum/1/ computation of design acceleration 5%-damping elastic spectra according to Eurocode 8 (EC8) (CEN 2003), Italian seismic code (NTC08) (CS.LL.PP. 2008) or user-defined spectra; this service is accessible to registered users.

4) https://tex.mi.ingv.it/esmws/rexel-spectra-selection/1/ provides machine-friendly access to response spectra included in the ESM Database; this service is accessible to registered users.

5) https://tex.mi.ingv.it/esmws/rexel-run-matching/1/ searches for sets of waveforms, whose spectra, on average, match the target spectrum within a specified tolerance; this service is accessible to registered users.

6) https://tex.mi.ingv.it/fdsnws/event/1/ provides machine-friendly access event-related metadata included in the ESM Database, following the standard of International Federation of Digital Seismograph Networks (FDSN); this service is accessible to registered users.

Additional tools are linked to the database, such as an interface to waveform processing (http://esm.mi.ingv.it/processing/) and a web interface to select a suite of 7 accelerograms compatible with the spectral shapes of the Eurocode 8 or the Italian seismic code (accessible from the database homepage). A parametric flat-file is released with annual rate, for engineering seismology studies (http://esm.mi.ingv.it/flatfile-2018/). Users must register to download or process the waveforms or download the parametric flat-file.

The version 2.0 of the Engineering Strong-Motion database (ESM) is ready in the alpha version and will be released in April 2020. The ESM database structure has been modified to archive supplementary information. The new database management system is PostgreSQL, while the back-end software is developed in Python. The graphic interface has been renewed (Fig. 3) to better support users' needs both in terms of contents and speed performance and the three main sections related to waveforms, events, and stations have been expanded to introduce additional information.

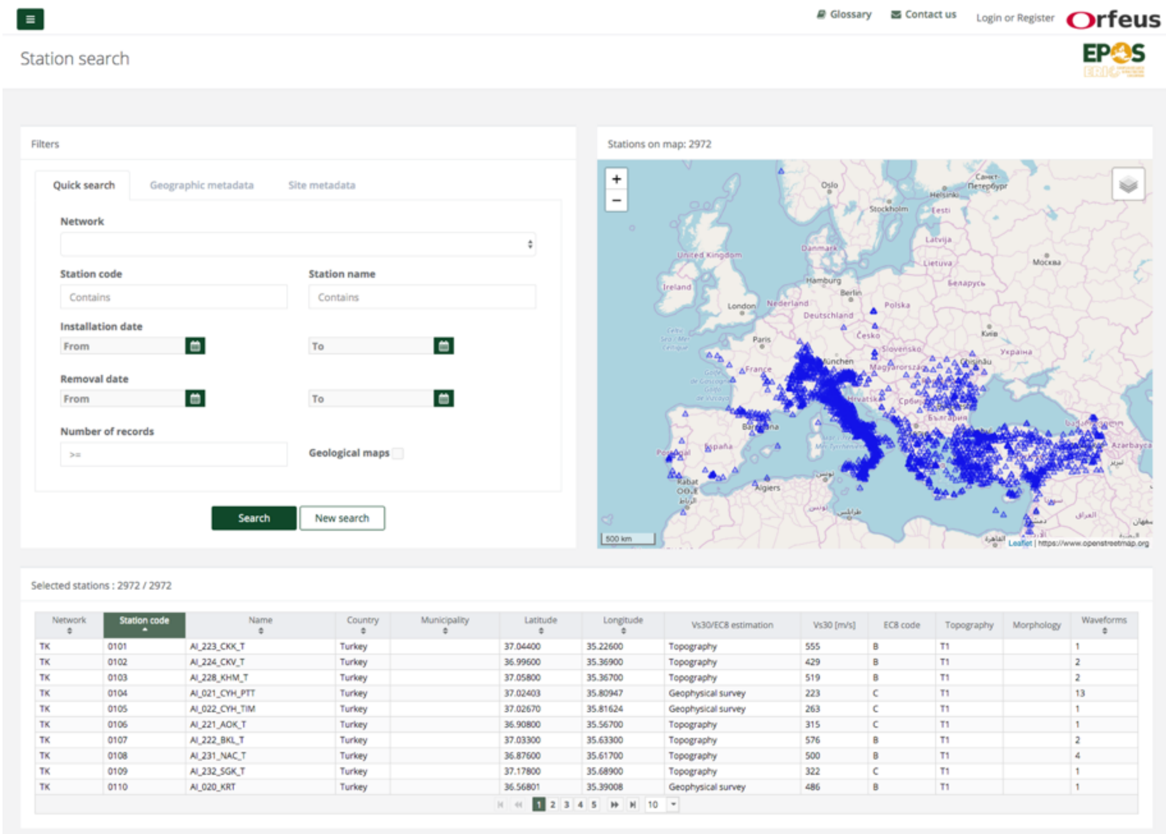


Figure 3: new web interface of the Engineering Strong-Motion database

1.3 European Archive of Historical Earthquake Data (AHEAD)

AHEAD, the European Archive of Historical Earthquake Data (1000-1899), is a distributed archive aiming at preserving, inventorying and providing, to investigators and other users, data sources on the earthquake history of Europe, such as papers, reports, Macroseismic Data Points (MDPs) or parametric catalogues. AHEAD consists of independent regional archives, a general repository and a collaborative inventory. At present, it mainly relies on eight regional, online macroseismic archives, which supply most of the data. It contains information on about 5000 earthquakes in the time-period 1000-1899 and provides parametric and macroseismic intensity data, derived from different sources, such as regional databases, papers, and catalogues. AHEAD establishes relationships among earthquake data of different provenance and provides multiple macroseismic intensity datasets and interpretations for each earthquake. About 230 data sources (papers, reports, catalogues) are available to users. AHEAD is the EPOS node for distributing historical earthquake data within EPOS.

The archive is accessible at https://www.emidius.eu/AHEAD with a user-friendly web interface. The archive can be queried by earthquake, by data source or by web services.

The query by event (https://www.emidius.eu/AHEAD/query_event/; see Fig. 4) allows to access the information related to each individual earthquake; the query by data source (https://www.emidius.eu/AHEAD/query_study/) allows the user to browse the list of the main studies available for the inventoried earthquakes and select them individually.

The query by web services (https://www.emidius.eu/AHEAD/services/, see Fig. 5) include: i) event parameters (FDSN-event, with a user-friendly query builder); ii) OGC (Open GeoSpatial Consortium): WFS (Web Feature Service) and WMS (Web Map Service); iii) Macroseismic intensity data and iv) Bibliographical metadata.

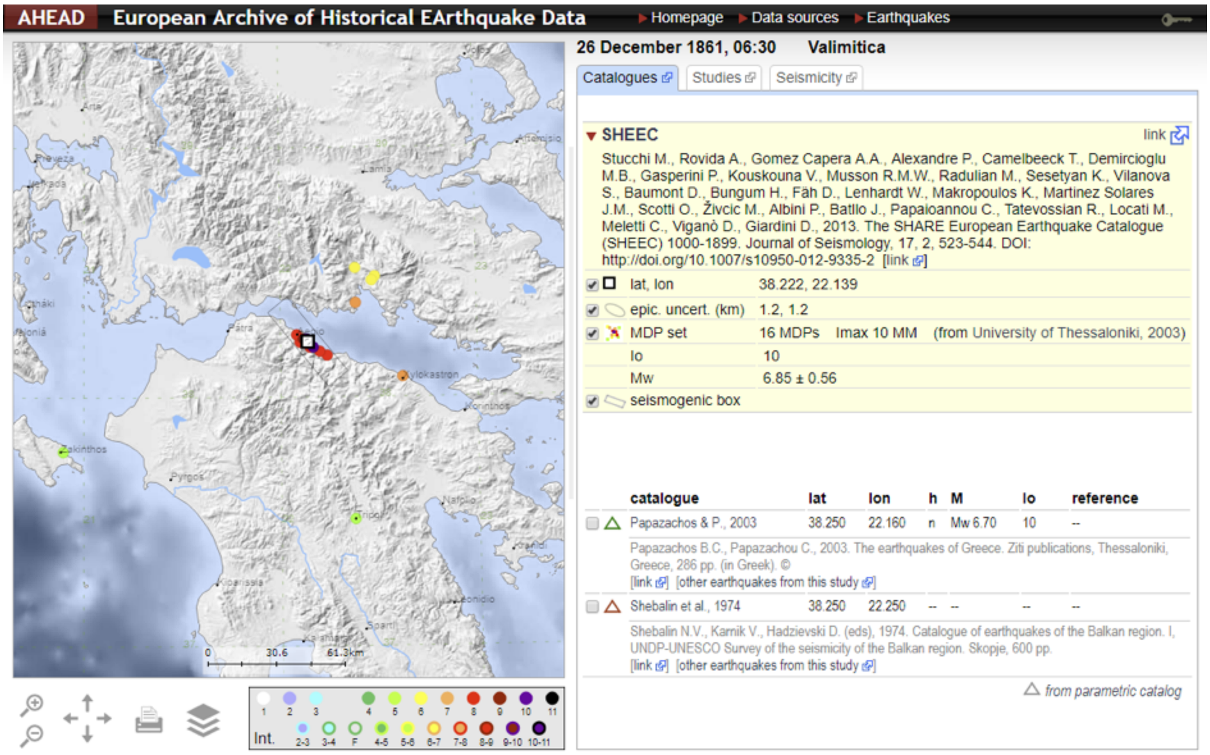
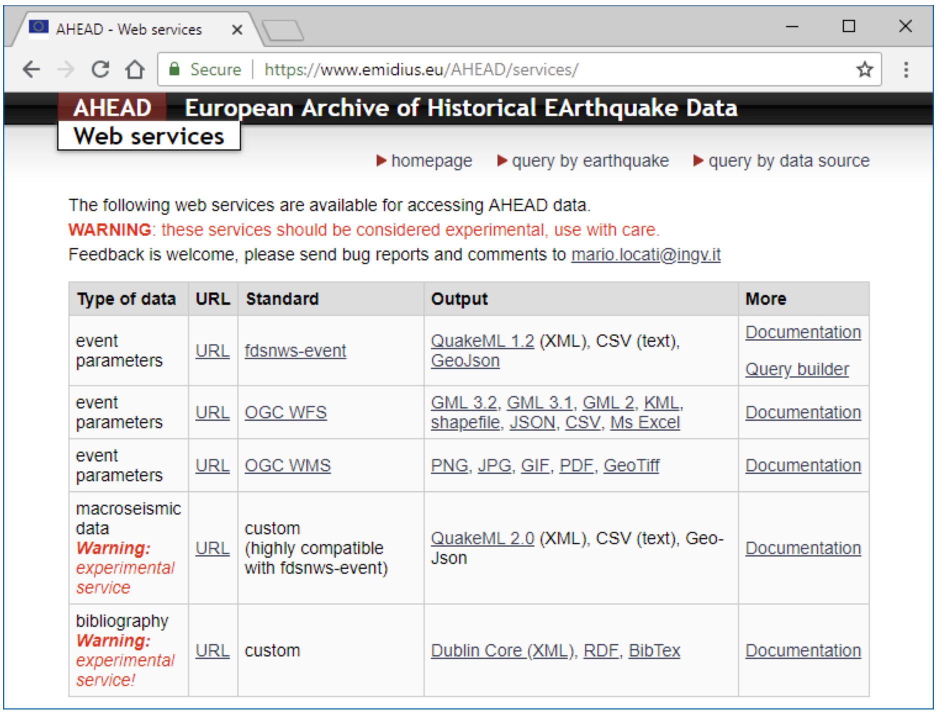


Figure 4: Example of a single event exploration.



Type of data	URL	Standard	Output	More
event parameters	URL	fdsnws-event	QuakeML 1.2 (XML), CSV (text), GeoJson	Documentation Query builder
event parameters	URL	OGC WFS	GML 3.2, GML 3.1, GML 2, KML, shapefile, JSON, CSV, Ms Excel	Documentation
event parameters	URL	OGC WMS	PNG, JPG, GIF, PDF, GeoTiff	Documentation
macroseismic data *Warning: experimental service*	URL	custom (highly compatible with fdsnws-event)	QuakeML 2.0 (XML), CSV (text), Geo-Json	Documentation
bibliography *Warning: experimental service!*	URL	custom	Dublin Core (XML), RDF, BibTex	Documentation

Figure 5: Web services available at AHEAD

1.4 European Database of Seismogenic Faults (EDSF)

EDSF collects and grants access to data on seismogenic faults of the Euro-Mediterranean region. It deals with 1128 crustal faults (for a total length of ~63,775 km) and 3 subduction zones (Calabrian Arc, Hellenic Arc, and Cyprus Arc; all located in the eastern Mediterranean region). EDSF provides parametric information and contextual references on geometry and behavior of potential seismogenic faults deemed capable of generating earthquakes of magnitude ≥ 5.5. Their identification and characterization were based on papers, original data, and empirical/analytical relationships. More than one hundred scientists from several pan-European institutions contributed to the development of the database in the framework of the EU-FP7 project SHARE. The original database is being distributed through the website http://diss.rm.ingv.it/share-edsf/ since February 2013 (Fig. 6), where it can be accessed through a user-friendly web interface, including a map viewer linked to parametric descriptions and references (Fig. 7).

As of today, EDSF is the node for distributing seismogenic fault data within EPOS-Seismology (https://www.epos-ip.org/tcs/seismology), through the European Facilities for Earthquake Hazard & Risk (EFEHR; http://www.efehr.org/). A new web portal has been developed (http://www.seismofaults.eu/; Fig. 8) providing access to the original platform, as well as to the newly implemented web services following the standard protocols of the Open GeoSpatial Consortium (OGC, http://www.opengeospatial.org/about). The following services have been implemented so far:

1. The WMS catalogue includes:
    a. GetCapability call to the collection of layers;
    b. Crustal Fault planes;
    c. Crustal Fault top;
    d. Subduction Zones;
    e. Subduction Contours;
    f. Crustal Faults (planes + top);
    g. Subduction Areas (zones + contour); and
    h. EDSF whole dataset.
2. The WFS catalogue includes:
    a. GetCapability call to the collection of layers;
    b. Crustal Fault planes;
    c. Crustal Fault top;
    d. Subduction Zones;
    e. Subduction Contours.

Figure 6: EDSF web site.

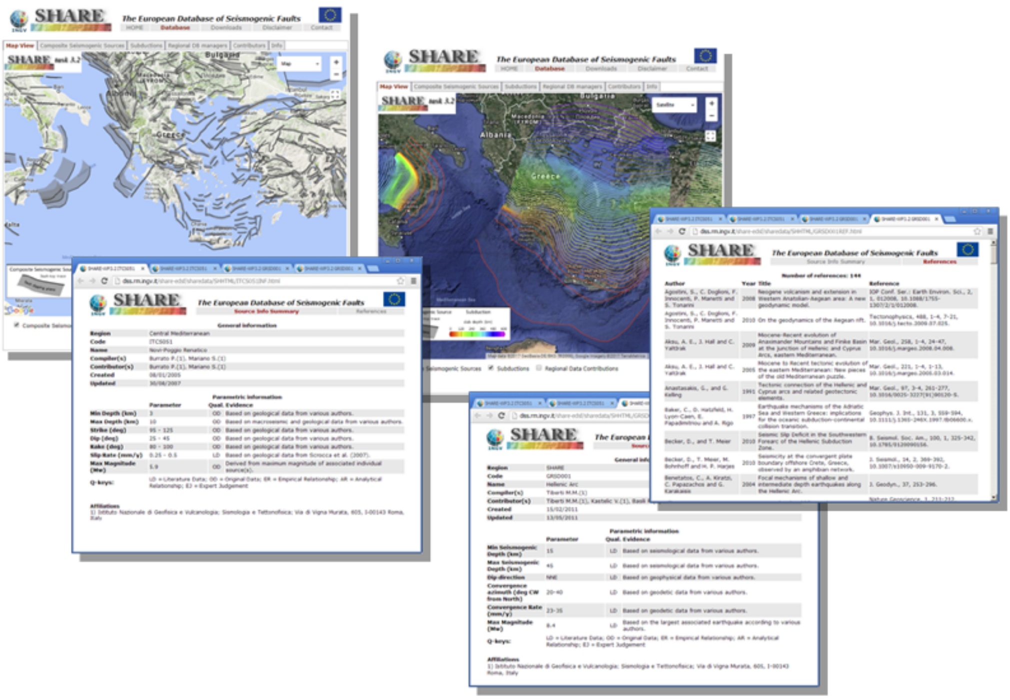


Figure 7: Navigation example of EDSF records from the original website.

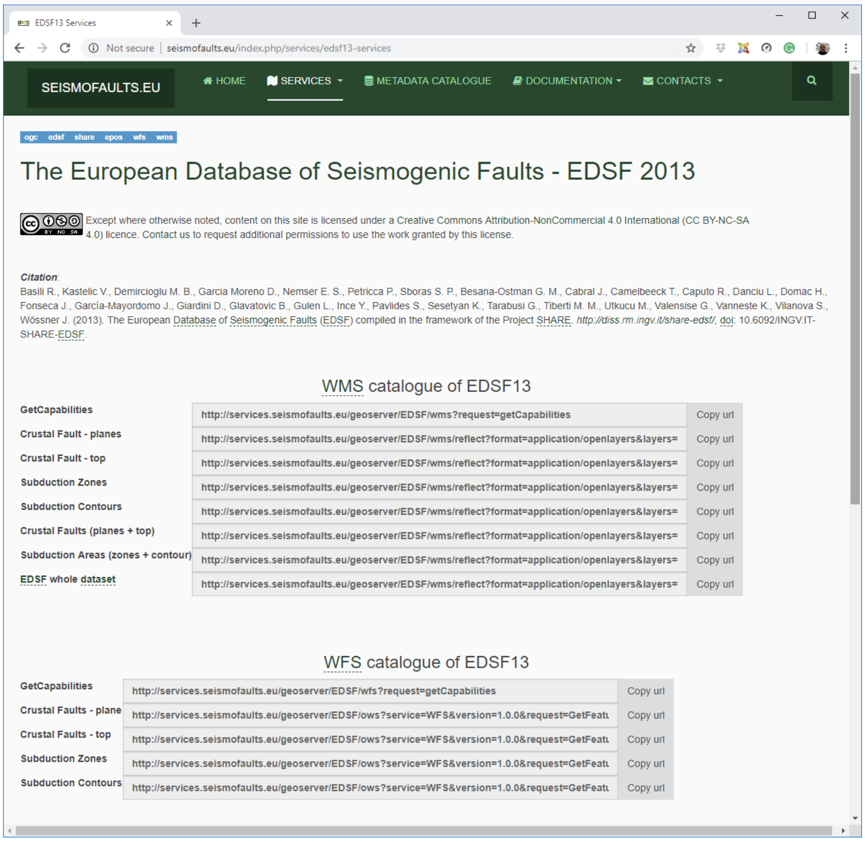


Figure 8: Web access to EDSF OGC services.

2 Chapter 2: Access statistic

The access statistics to the VA3 are generated using AWStats (https://awstats.sourceforge.io/), a free, powerful and highly customizable tool distributed under the GNU General Public License, that generates advanced web, streaming, ftp or mail server statistics.

The following sections illustrate the access distributions for the VA3 portal, and the three services described in Chapter 1. Since they run on independent web servers and distribute different types of data (e.g. waveforms, earthquakes parameters, macroseismic points, seismogenic faults), the access statistics for each service is shown separately. The reference period for this deliverable is from May 2017 (beginning of the project) and July 2018.

2.1 VA3 web portal statistics

At the time of the previous reporting phase (data until July 2018; Fig. 9) the VA3 web portal was a very recent product, as such it was not known within the users' community and most of its functionalities were still under development. Usage statistics (Fig. 10) for this reporting phase reflects the navigation behavior of the targeted users.

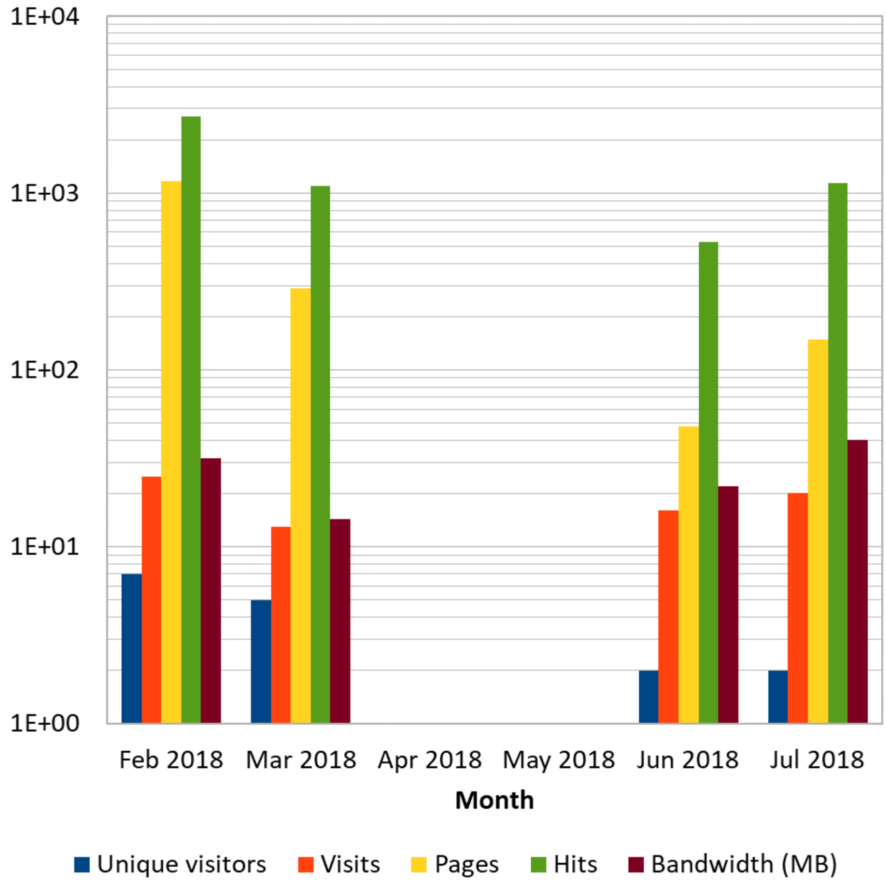


Figure 9: VA3 portal website usage related to the first reporting period: data are limited to periods of first implementation and testing.

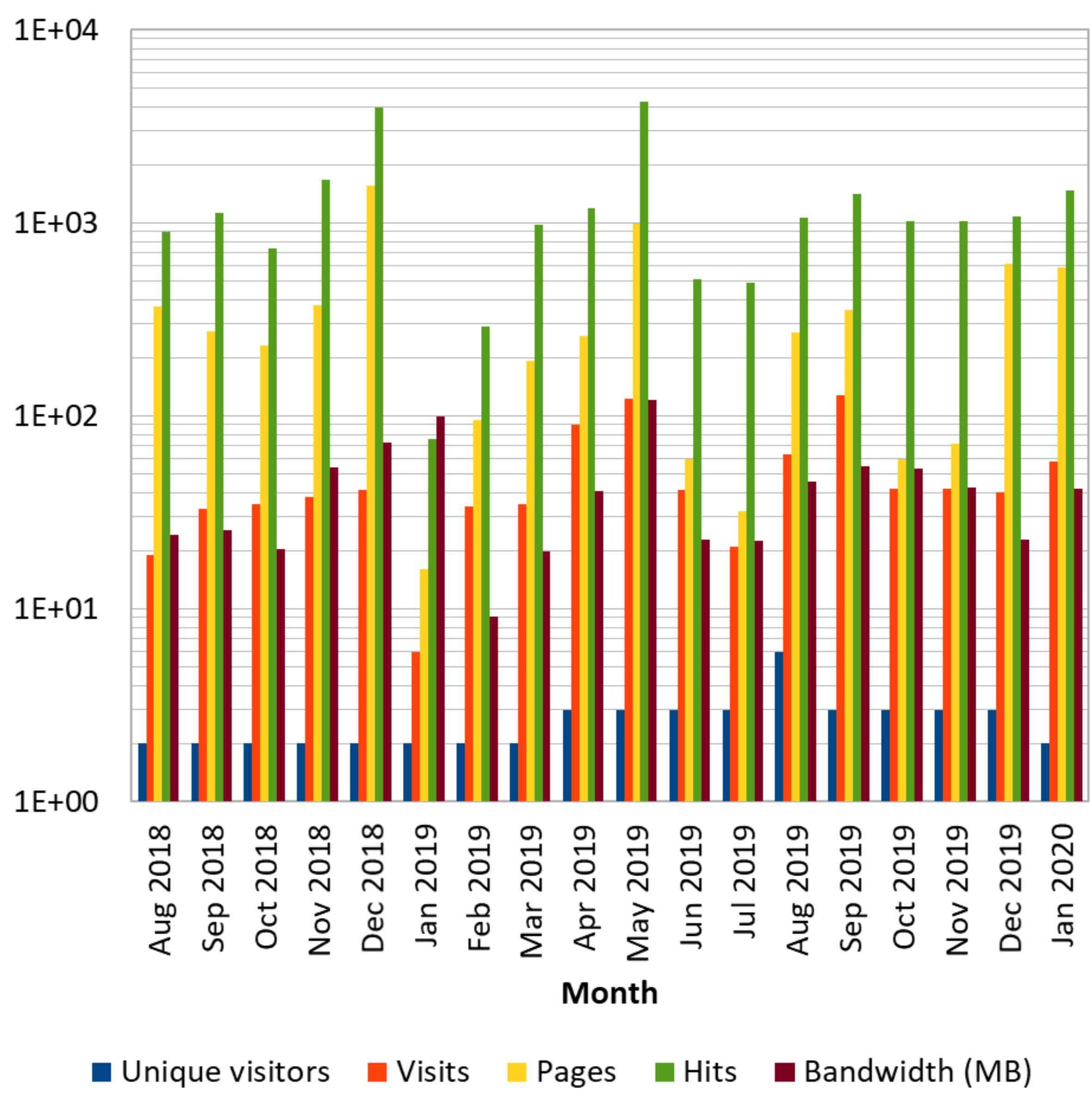


Figure 10: VA3 portal usage in current reporting period.

2.2 ESM statistics

Fig. 11 shows the number of unique visitors of ESM per month, which is generally larger than 1000 unit. Fig. 12 shows the number of repeated visits per month, which generally doubles the number of unique visits; the number of visited pages varies from 50k to nearly 500k per month (Fig. 13), which implies 25 to 50 pages per visitor. The employed bandwidth ranges from 5 to 70 GB (Fig. 14). An evident increase in the number of visited pages and used bandwidth can be attributable to the Albanian earthquake (26 November 2019). In fact, there are about 500k pages visited in November 2019 and more than 60Gb downloaded during November 2019 to January 2020.

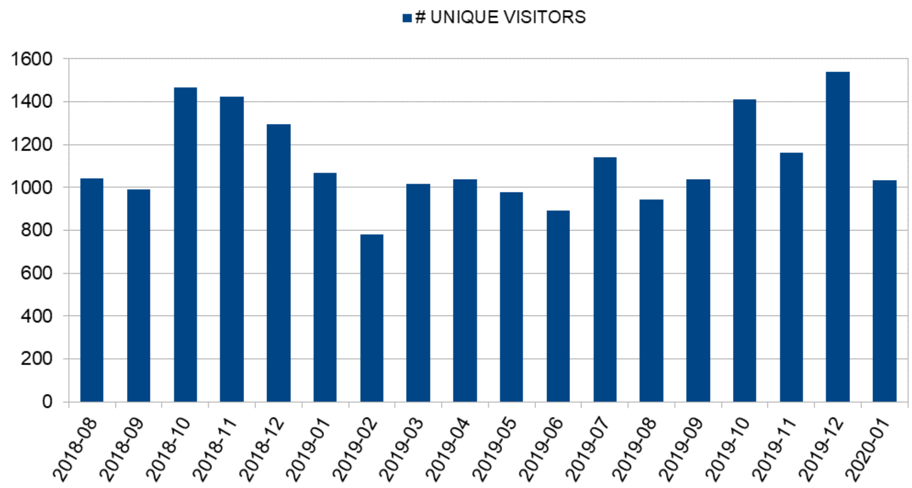


Figure 11: ESM unique visitors

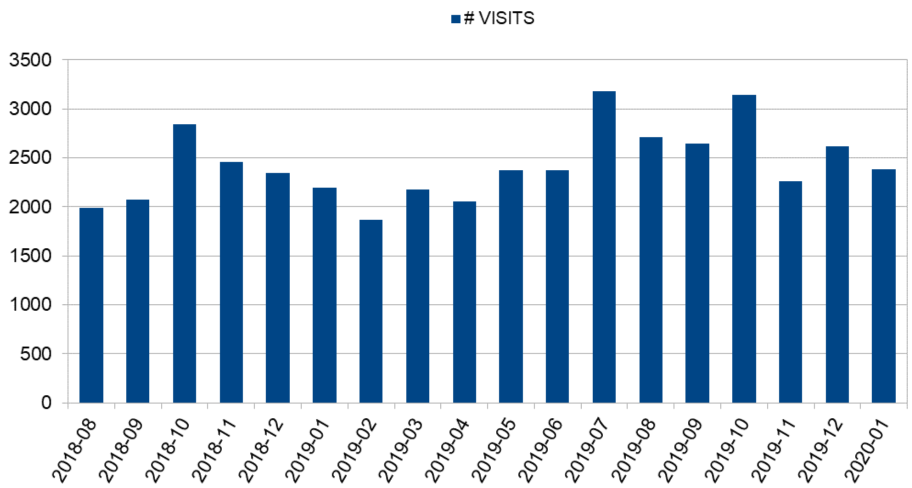


Figure 12: ESM number of visits

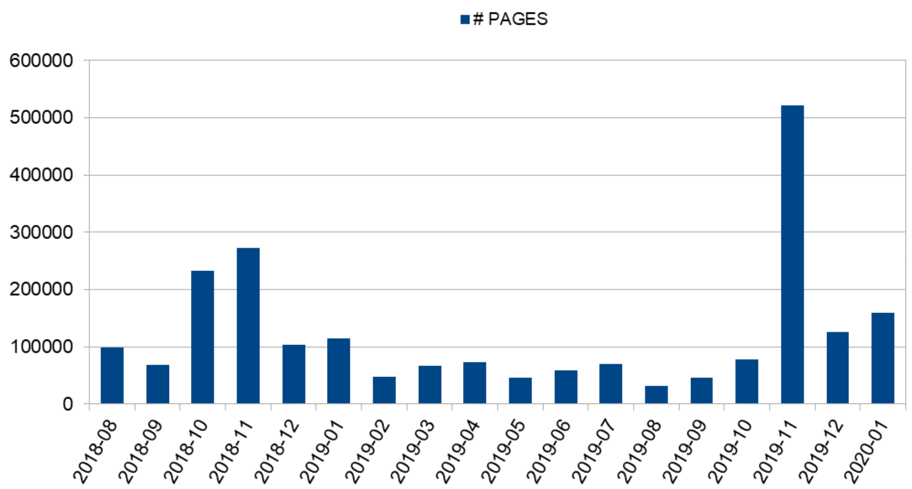


Figure 13: ESM number of visited pages

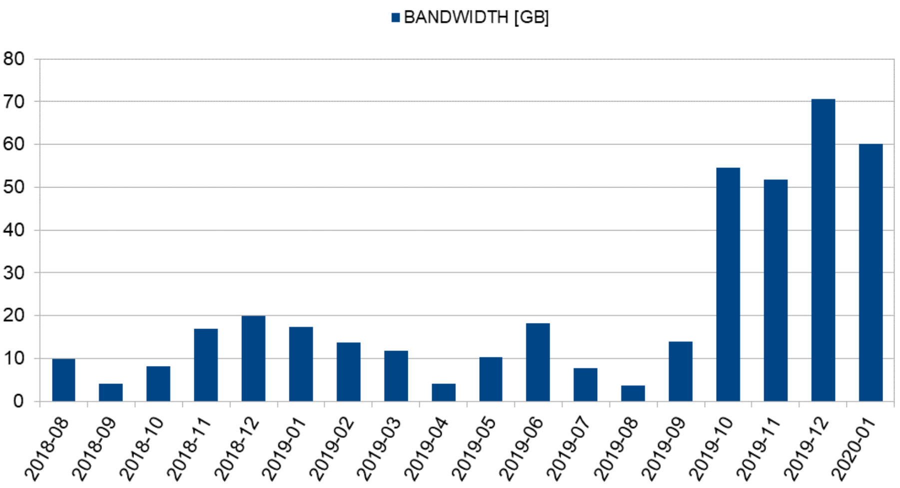


Figure 14: ESM bandwidth per month

When the ESM users are analyzed and compared to the users in the period May 2017- July 2018, it is evident the increase of visibility outside Italy. In fact, in the second period about 30% of the pages have been visited by Italians, while the percentage was about 50% in the period May 2017- July 2018 (Fig. 15). The remaining 70% of users are from European countries or mainly from US, Cameroon, China, and Japan, outside Europe. There is also a large number of users whose countries could not be identified (21%). The same statistics holds for the bandwidth (Fig. 16).

ESM has more than 2000 registered users, with an increase of 1000 units in the period August 2018-January 2020. The large majority comes from the academic world of from public administrations, while a minority comes from private companies or from the consulting/freelance world (Fig. 17).

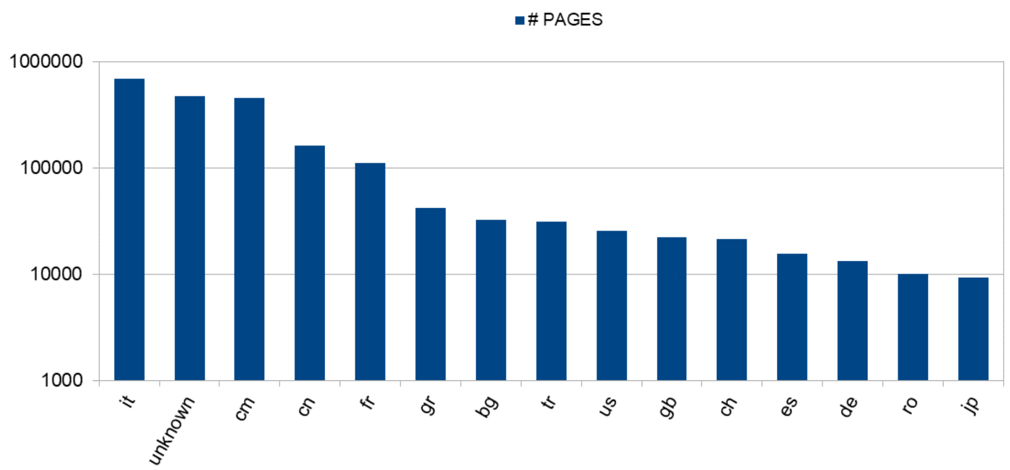


Figure 15: ESM visited pages per country

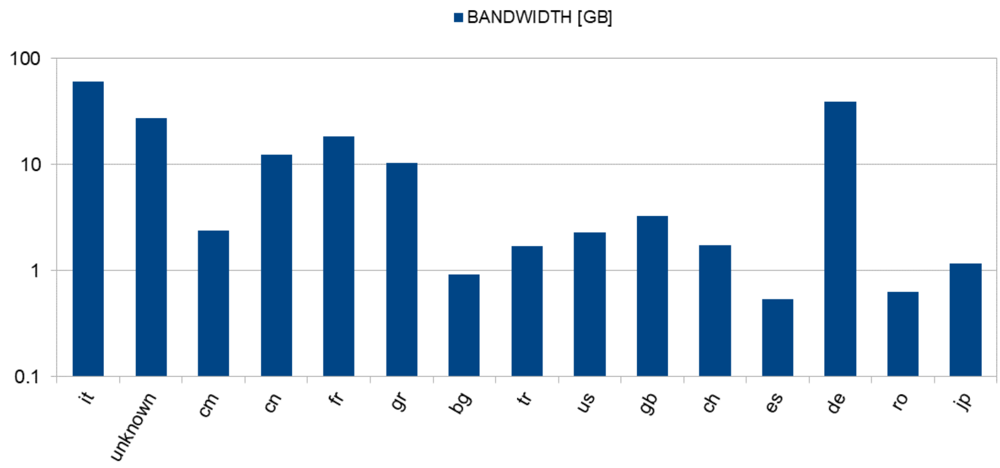


Figure 16: ESM bandwidth per country

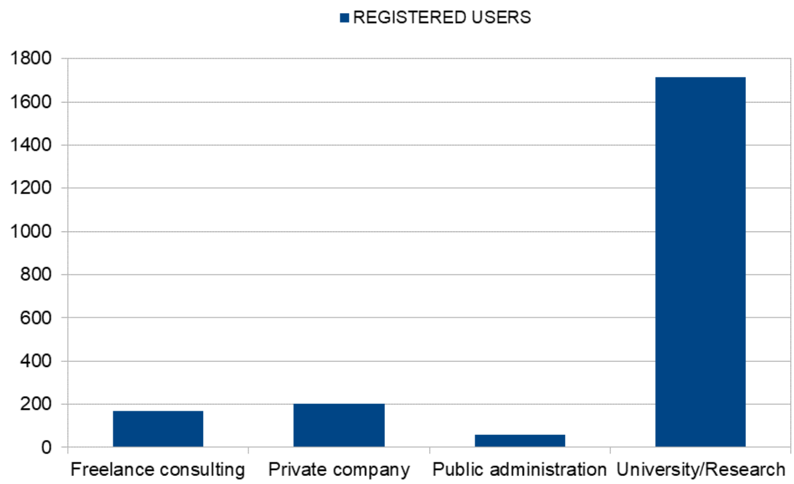


Figure 17: ESM type of users

2.3 AHEAD statistics

As it was described in the previous report, AHEAD data may be accessed either via traditional web pages, as well as via a variety of web services, listed at https://www.emidius.eu/AHEAD/services/ .

In the considered period of time of this report, August 2019 and January 2020, the number of unique visitors also drastically increased, from a 500 to 800 users per month, in the previously reported period May 2017 - July 2018, up to 600 to 1800 users per month (Fig. 18).

The use of the AHEAD web services drastically increased as well. In most cases, we see that web services outperform the user access via traditional web pages, demonstrating that users greatly appreciate their convenience. The analysis of the used bandwidth (Fig. 21) clearly reflect that today most of the data is transferred using web services.

The most used web services (Table 1), still is the OGC WMS (45.3% of all requests), most probably because it can be easily incorporated in any OGC compliant application, either a standalone software such as GISs, or within a web applications via widespread geographical APIs (i.e. leaflet). Another reason that could possibly explain its use, is that it provides already styled maps (i.e. represented epicenters does comes with a specific rule-based style), whereas all other services provide raw geographical features with no particular representation style associated. Finally, an additional reason could be possibly linked to the inclusion of the AHEAD based WMS service in the EPOS ICS data portal (https://ics-c.epos-ip.org/data/search). The second most used web service is the FDSN-event, quite used by seismologists as it allows to fine tune its outputs thanks to the free combination of useful query parameters.

By far, the vast majority of requests to AHEAD web pages and services still comes Italy (Fig. 22), followed by other European countries. However, when considering the used bandwidth for transferring data (Fig. 23), the United States comes at second place, well above the average use of other European countries.

The top 10 most requested earthquakes data and the top 10 most requested data sources are shown respectively in Table 2 and Table 3.

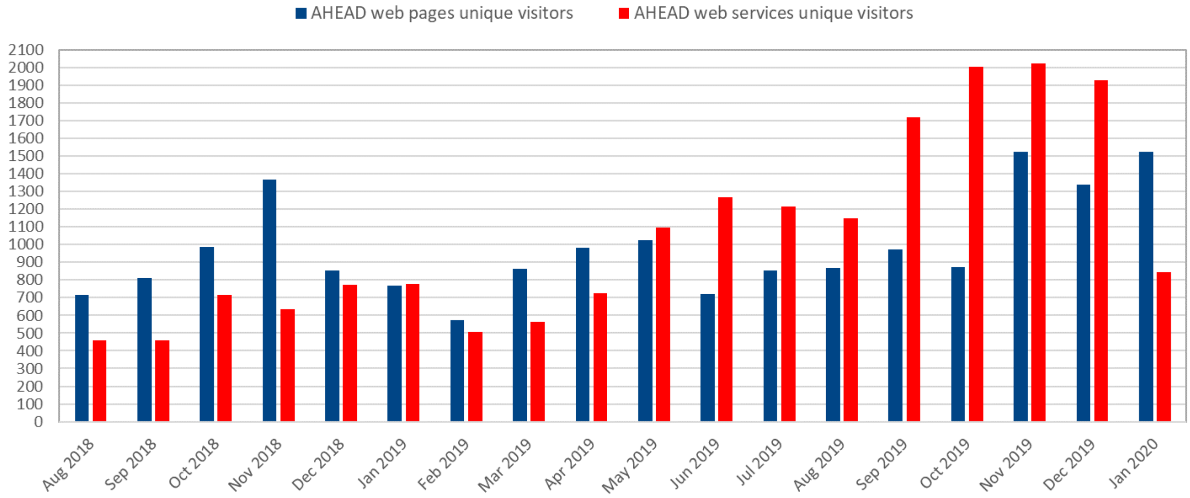


Figure 18 AHEAD number of unique visitors.

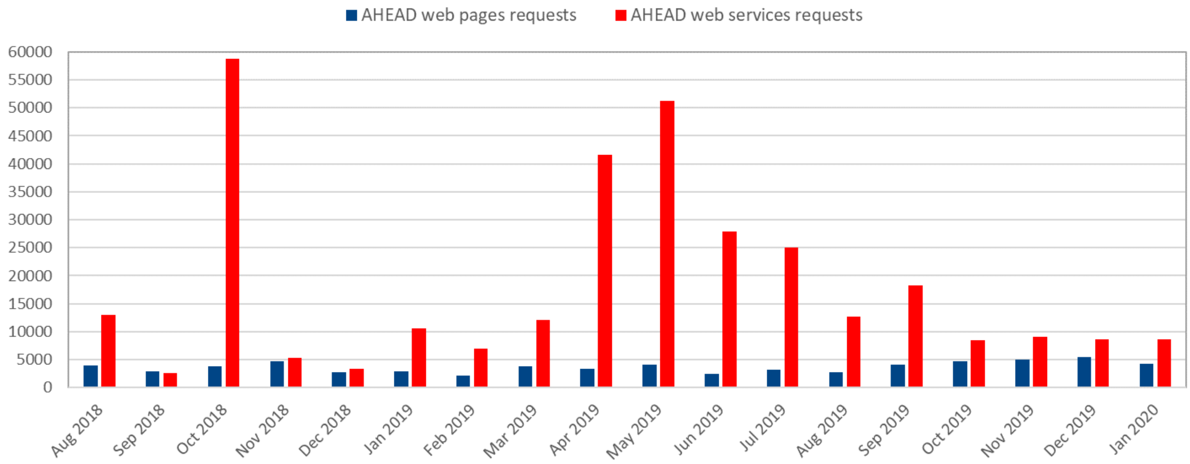


Figure 19: AHEAD number of requested pages.

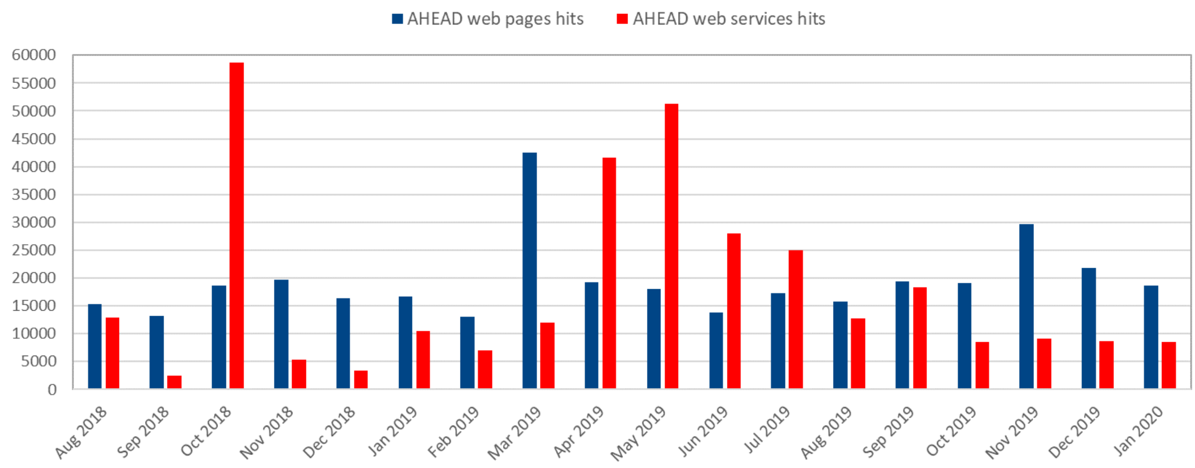


Figure 20: AHEAD number of hits.

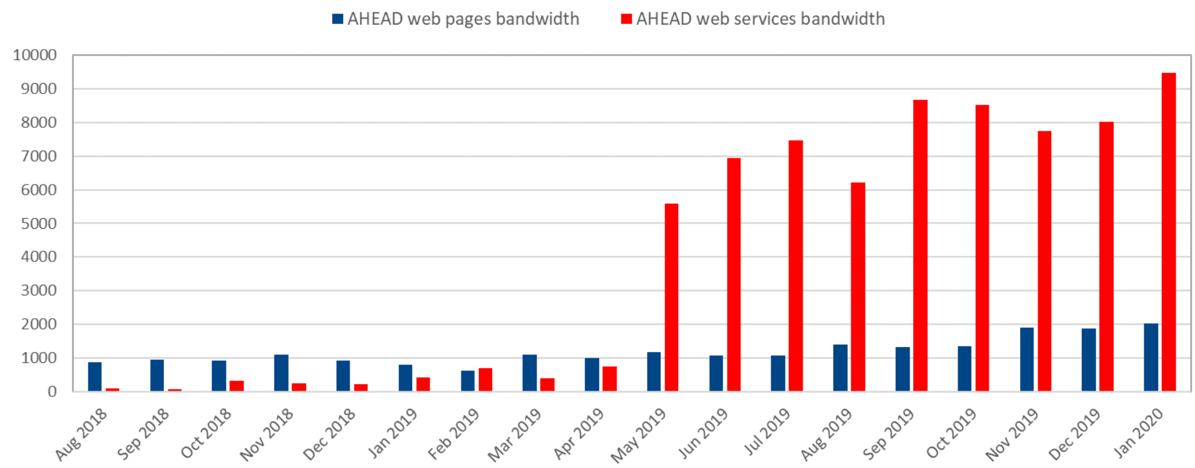


Figure 21: AHEAD bandwidth usage.

TYPE OF WEB SERVICE	REQUESTS	% REQUESTS
OGC WMS	32352	45.27
FDSNWS-EVENT	18546	25.95
MACROSEISMIC	8429	11.79
OGC WFS	7647	10.70
BIBLIOGRAPHY	4492	6.29

Table 1: AHEAD number of requests by type of Web Service.

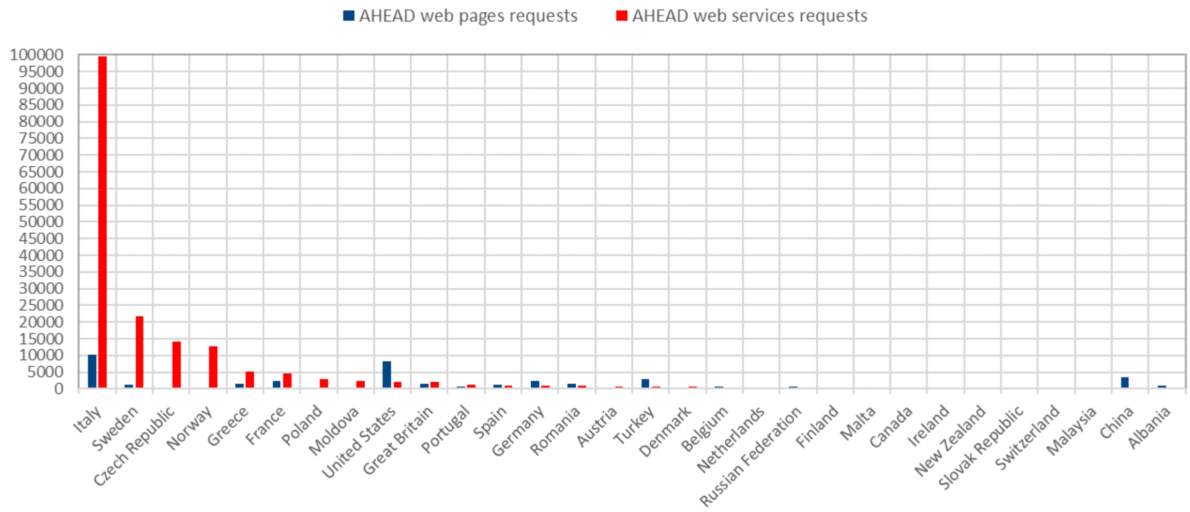


Figure 22: AHEAD pages requested per country (Top 30).

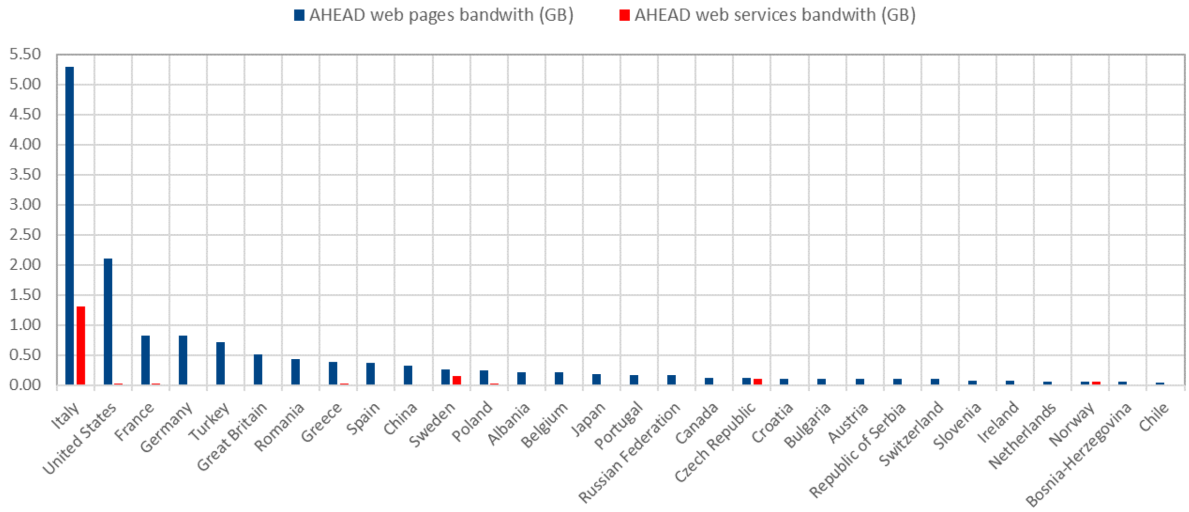


Figure 23: AHEAD bandwidth per country (Top 30).

EVENTID	DATE	EPICENTRAL AREA	REQUESTS
13480125_1530_000	1348 01 25 15 30	Carinzia	158
18611226_0630_000	1861 12 26 06 30	Valimitica	132
16880605_1530_000	1688 06 05 15 30	Sannio	128
17551101_0930_000	1755 11 01 09 30	Lisboa	124
14561205_0000_000	1456 12 05	Molise	107
16901204_1400_000	1690 12 04 14	Carinzia	99
16930111_1330_000	1693 01 11 13 30	Sicilia orientale	94
18960210_0056_000	1896 02 10 00 56	[Durres]	88
18870717_0745_000	1887 07 17 07 45	Heraklio	88
14710829_0800_000	1471 08 29 08	Vrancea	86

Table 2: AHEAD Top 10 most viewed earthquakes.

DATA SOURCE	REQUESTS
Papazachos and Papazachou, 2003 *The earthquakes of Greece*	226
Ambraseys and Finkel, 1995 *The Seismicity of Turkey and adjacent areas. A Historical Review, 1500-1800*	121
CPTI04 *Catalogo Parametrico dei Terremoti Italiani, versione 2004 (CPTI04)*	115

Boschi and Guidoboni, 2001 *Catania terremoti e lave dal mondo antico alla fine del Novecento*	112
Ahjos and Uski, 1992 *Earthquakes in Northern Europe in 1375-1989*	109
Soysal et al., 1981 *Turkye ve Cevresinin Tarihsel deprem Katalogu*	104
Ambraseys, 2002 *The seismic activity of the Marmara Sea Region over the last 2000 years*	98
Guidoboni and Comastri, 2005 *Catalogue of earthquakes and tsunamis in the Mediterranean area from the 11th to the 15th century*	91
Ambraseys et al., 1994 *The Seismicity of Egypt, Arabia and the Red Sea*	86
Martinez Solares and Mezcua Rodriguez, 2002 *Catalogo sismico de la Peninsula Iberica (880 a.C.-1990)*	79

Table 3: AHEAD Top 10 most viewed data sources (literature).

2.4 EDSF statistics

EDSF usage statistics are subdivided into two categories: 1) web pages, and 2) web services. Web pages regard only the visits to the website http://diss.rm.ingv.it/share-edsf which introduces to the outcomes of project SHARE in terms of distributing seismogenic fault information of the Euro-Mediterranean region. The use of the data though the original EDSF website, including visits to the map viewer, connected pages, and file downloads, are only counted here. Web services regard the use of the data distributed via OGC standard protocols (see also the Seismofaults website http://www.seismofaults.eu/), excluding the just mentioned categories of information.

For monitoring the website and the web services we used either the Google Analytics online tools or the software AWStats.

The server that distributes EDSF website and web services is hosted in an area of the INGV data center of Rome that for both security and legal issues filters the connection through a proxy system. This implies that it was not possible to identify the source IP connection.

Since the end of October 2019, our server was provided with a public IP address, and since then it became possible to track the source IP of clients that connect to Seismofaults/EDSF web pages and services. This means that we are now able to provide a more detailed statistics including the number of "unique users" who rely on our services. However, because of the short period for which such data where registered, we considered it of little significance for the time being.

Fig. 20 shows accesses to the EDSF website from August 2018 to January 2020. The distribution of "sessions" appears to be rather uniform during most of the observation period. During the last three months there has been a significant increase of the visits, probably as a response to the release of the stable version of EPOS ICS-C web portal (https://ics-c.epos-ip.org).

Instead no significant changes where recorded in the number of visited pages or in the number of "data pages" downloaded (Fig. 25).

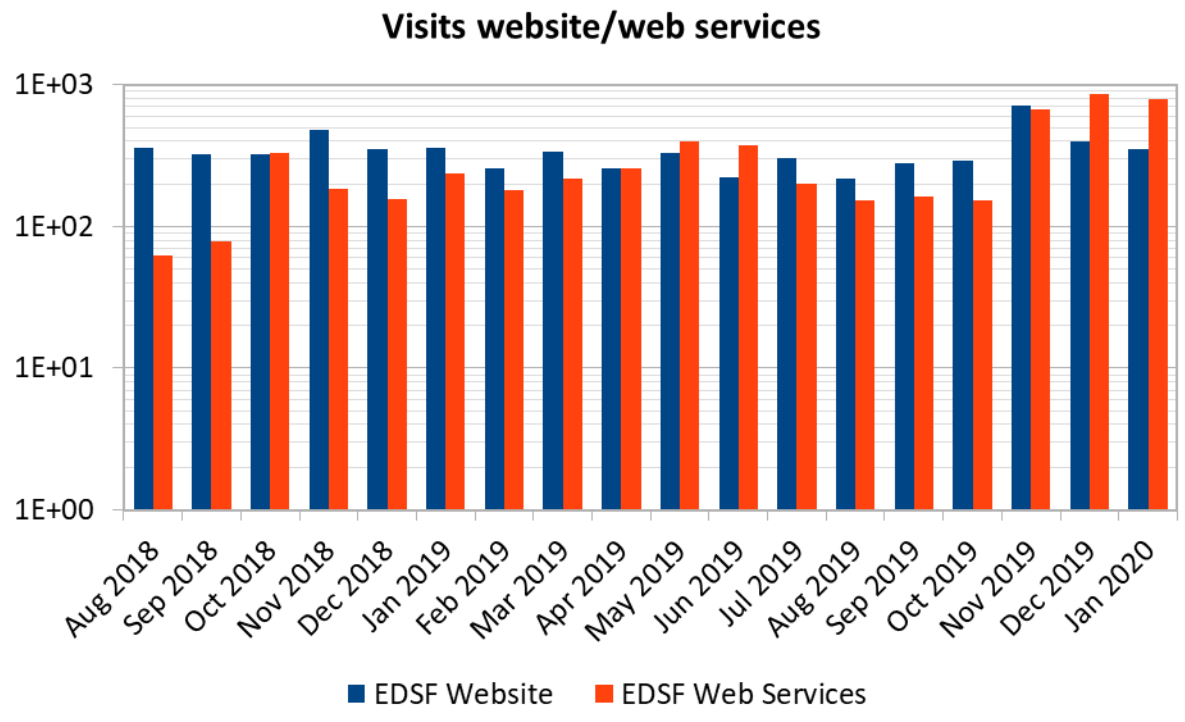


Figure 24: EDSF number of visits.

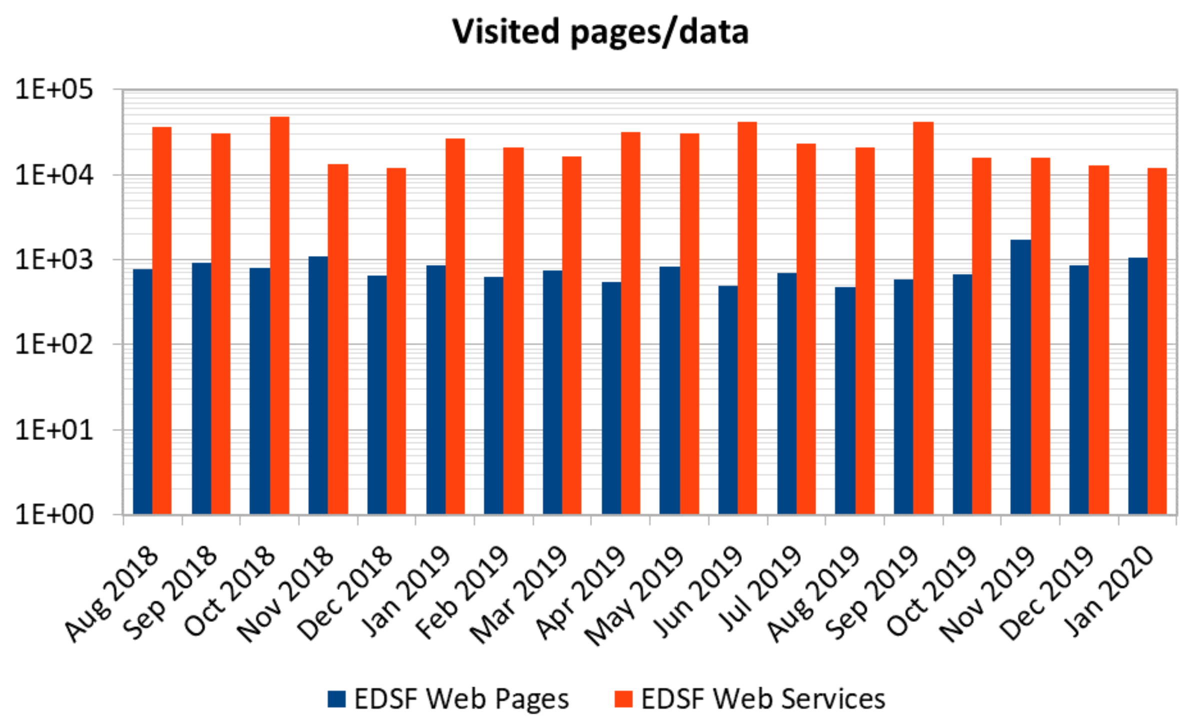


Figure 25: EDSF number of visited pages.

In Table 4 the use of web services is expressed by distinguishing between per-service and per-layer ways. The request breakdown by category: WMS, WFS, and OWS are also shown.

Table 4: EDSF visited page breakdown by type and web service.

TYPE OF WEB SERVICE	CRUSTAL FAULT	SUBDUCTION	UNDEFINED	TOT
OGC WMS	14110	13642	14509	42261
OGC WFS	5985	5999	7495	19479
OGC OWS	0	0	184381	184381

Compared to the last reporting period, an important increase of the requests classified as “undefined” as well as the requests made through the generic “OWS” service can be observed. This is likely due to the gradual increase of the EPOS ICS-C web portal usage. In the EPOS platform, the EDSF web services were mapped through a specific EPOS-DCAT-AP metadata (https://github.com/epos-eu/EPOS-DCAT-AP) in which they are linked to a URI that includes “OWS”.

The “Undefined” data refer to generic requests to the entire dataset made by the clients. This kind of requests started increasing in 2019 and progressively replacing the per-layer requests as shown in Table 4. This is likely also due to the EPOS ICS-C platform in which the entire EDSF dataset is loaded by default.

Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D20.1 Report on access statistics and service provision of VA1-VA5

(PART 4/ETH)

Work package	WP21 – Virtual Access VA4: European Facilities for Earthquake Hazard and Risk
Lead	ETH Zurich, Switzerland
Authors	L. Danciu, P. Kästli, Swiss Seismological Service, ETH Zurich, Switzerland
Reviewers	J. Douglas, University of Strathclyde, UK, R. Bossu (EMSC), K. Saleh (ETH)
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	29.02.2020
Submission date	30.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D20.1_Report_Virtual_Access_statistics.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

This document provides an overview of the main services of the European Facilities of Earthquake Hazard and Risk (EFEHR – www.efehr.org) developed and maintained under SERA Virtual Access (VA) WP2a. This report provides insights on the usability of these services in the reporting period March 2019 to March 2020.

The web-traffic analytics of the EFEHR web-portal indicates a preference for users to access the hazard maps and uniform hazard spectra. A total number of 58 000 unique visitors with a monthly average of about 2500 new users have been accessing and/or visualizing the hazard data, models and products (hazard maps, curves and spectra).

The visitors are distributed worldwide, and top ten countries to access the EFEHR resources in the reporting period, in Europe are: Switzerland, Italy, United Kingdom, Greece, Germany, Turkey, France, Spain, Albania and Netherlands. Outside the European region, the top ten countries ranked in terms of terms of users accessing or visualizing the EFEHR resources are: United States, Iran, China, South Korea, Canada, Russian Federation, India, Egypt, Singapore and Japan. The ranking is a combine index of “http page hits”, web-page display and web traffic.

Often the visitors are consulting the hazard values at a specific site, rather than downloading entire sets of results and/or models. An average of 6000 requests per month are observed for all the web-services.

The most used web application is the hazard map viewer. In terms of outcomes resulting from this access, there are about 650 citations of the datasets, results, models provided by EFEHR web-portal since 2013. Note that this is an update of the SERA Deliverable 18.1, hence the technical description of the main services will not be repeated here.

1 Overview of the SERA-VA4: European Facilities for Earthquake Hazard and Risk – EFEHR

European Facilities for Earthquake Hazard and Risk (EFEHR) provides open access to existing seismic hazard models and results via a web-platform (www.efehr.org) hosted and maintained at ETH Zurich.

The portal consists of web-services linking to the main hazard outputs (i.e. seismic hazard maps, seismic hazard curves and uniform hazard spectra).

The webservices are built upon various technologies summarized in the Deliverable 18.1 and not repeated here. In terms of available hazard models, the EFEHR web-portal distributes the seismic hazard models for:

- The 2013 European Seismic Hazard Model (ESHM13, Wössner et al 2015)
- The 2014 Earthquake Model of the Middle East (EMME14, Giardini 2018)
- The 2015 Swiss Hazard Model (SuiHaz15, Wiemer et al 2015)
- The 1999 Global Hazard Map of the Global Seismic Hazard Assessment Program (GSHAP, Giardini 1999)
- The 2020 European Seismic Hazard Model (ESHM20, outcome of the SERA JRA3, to be updated in the next six months, after the model is finalized)

The EFEHR web-portal provides a single access point for data, models and results. No user authorization is required.

Update of Data Access Applications

The EFEHR web platform consists of three stand-alone web applications for interactively discovering and retrieving hazard curves (Figure 1, upper), and hazard spectra (Figure 1, lower) and hazard maps (Figure 2).

Two web applications (i.e. hazard spectra and hazard maps) have been updated during the last reporting period (March 2019 to March 2020).

The new layer manager of the Hazard Spectra web application is fully customizable allowing the addition of a hazard map, input datasets and/or control different transparency levels.

The background layers allow adding Open Source Map content, country boundaries and control the transparency of these layers.

Additional layers can be added to the map view to illustrate some of the main components of the hazard model: earthquake catalogue, active faults and the source zonation.

A log panel is located at the bottom of the web page. It provides a summary of the model query parameters and a URL link to download the selected data.

The hazard map viewer has been updated to a new interface and a layer manager (same as for spectra). The layout of the webpage has been preserved, with the model selection consisting of drop-down menus on the left hand-side.

The menu bar appears in all data viewers except the documentation section. Map Controls allows controlling the map viewer, and Print allows one to print the entire web-page and/or a selected area.

The Map Legend is located on the right-hand side of the webpage, indicating the color bar of the selected hazard maps. By default, the background layers are illustrating the country boundaries, the geo-reference of Europe from OpenStreet Maps and the transparency controllers of all layers

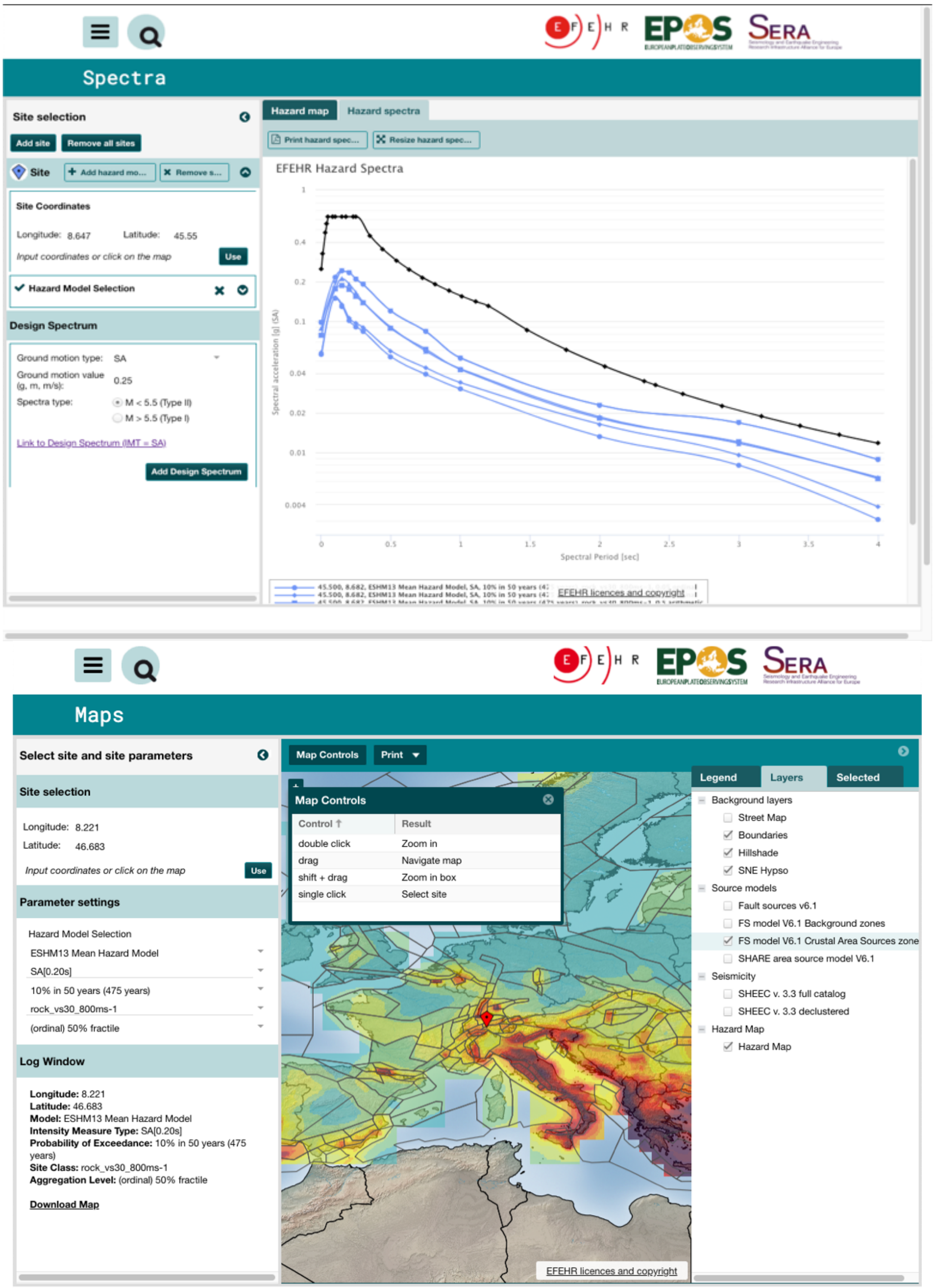


Figure 1: EFEHR web-portal: hazard spectra (mid) and hazard map viewer application (bottom)

2 EFEHR Web-portal Usage and Metrics

General user characteristics

In this report we summarize the statistics on the reporting period March 2019 to March 2020. Visitors of the EFEHR web platform come from all over the world. Most visitors of the EFEHR web platform already know about it beforehand:

46% the visits are either from bookmarks, directly by typing its web address or via a link in an email;

43% come from references (www.share-eu.org, www.seismicportal.eu, www.europe-geology.eu, www.link.springer.com, Wikipedia etc.); and

11% come from search engines.

A typical visitor (median) accesses 5 pages (either web pages or viewer applications). 15% of the visitors come back to the web portal within the same month. Also, most of users are either accessing the hazard at a specific site or just to view the maps, rather than downloading the data.

The average time spent by a user is about 3 min to download a data, 10% of users spent more than 15 minutes on the webpage, and 1% of web visits were there for more than one hour.

The number of unique visitors of EFEHR web-platform from May 2018 to April 2020 are illustrated in Figure 2. The total number of unique users is 58562 users with and average number of new users per month of about 2440.

An increased number of users is observed in June and July 2018 – this coincide with the release of the Earthquake Model of the Middle East (EMME14, Giardini et al 2018 has been released via the web-platform). Thus, the traffic increases with the release of a new model. Moreover, the increased number of unique users in November, might be linked to the occurrence of significant earthquakes in Europe i.e. the 5.4Mw in France (November 11th) and the 6.4Mw earthquake in Albania (November 24th).

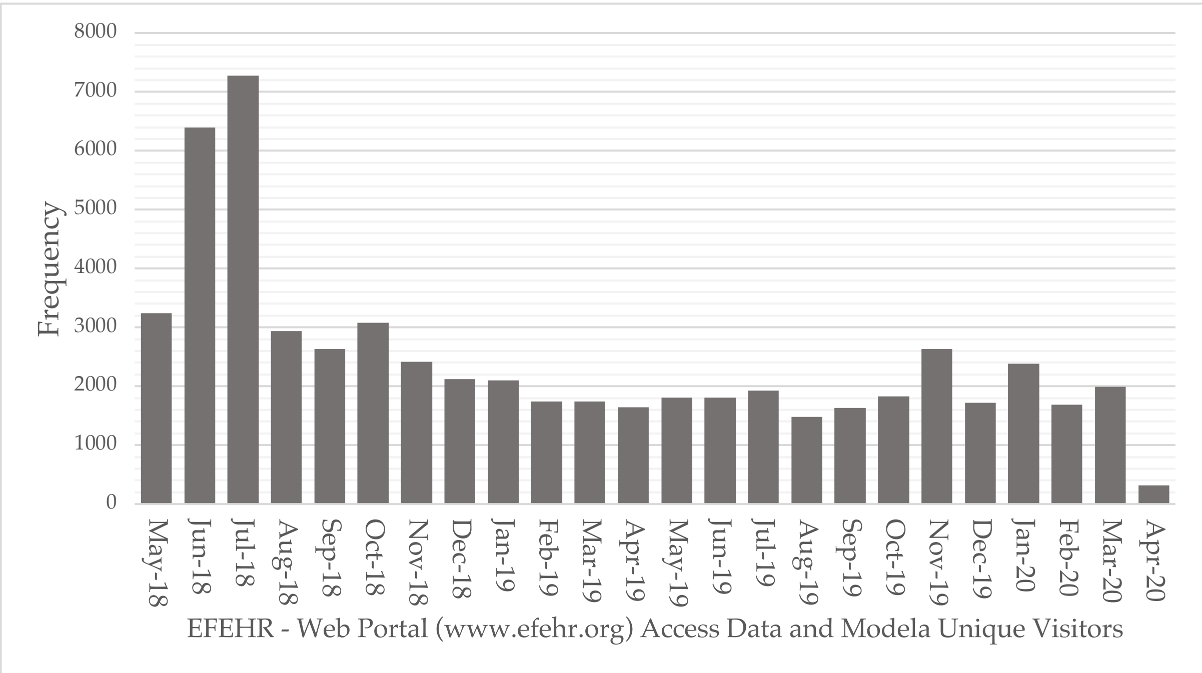


Figure 2: Frequency of new users accessing EFEHR web-portal in the reporting period May 2018 to April 2020

The web traffic of EFEHR is summarized in Figure 3 in terms of usage information defined as the number accessing the webpage, the number of page loads and the amount of data traffic in Gigabytes (Gb). The data underlined in these charts is given in Table 1.

Multiple consecutive page views within the same portal during the same day is quantified as one visit, hence the count accessing the web-portal is higher. Unfortunately, multiple page views within a single day cannot be monitored with the current implementation.

The spatial distribution of the page load per country within Europe is given in Figure 3. The top 10 countries in terms of accessing (joint access and traffic size) content on EFEHR web-platform are Switzerland, Italy, United Kingdom, Greece, Germany, Turkey, France, Spain, Albania and Netherlands.

The top 10 countries of EFEHR visitors outside the Pan-European region are: United States of America, Iran, China, South Korea, Canada, Russian Federation, India, Egypt, Japan and Lebanon. A summary of the access stats is illustrated in Figure 4 (top) for European Countries and Figure 4 (bottom) for non-European countries respectively.

The figures above correspond to "http page loads". Here, loading a data application counts as one-page load, independently of how many data, e.g. hazard curves, are downloaded afterwards. However, to reset the application after one single download, a user may either use the application's reset button or just reload the entire application. Inadequate scripting of interactive access may contribute to some peculiarities in numbers of requests per country.

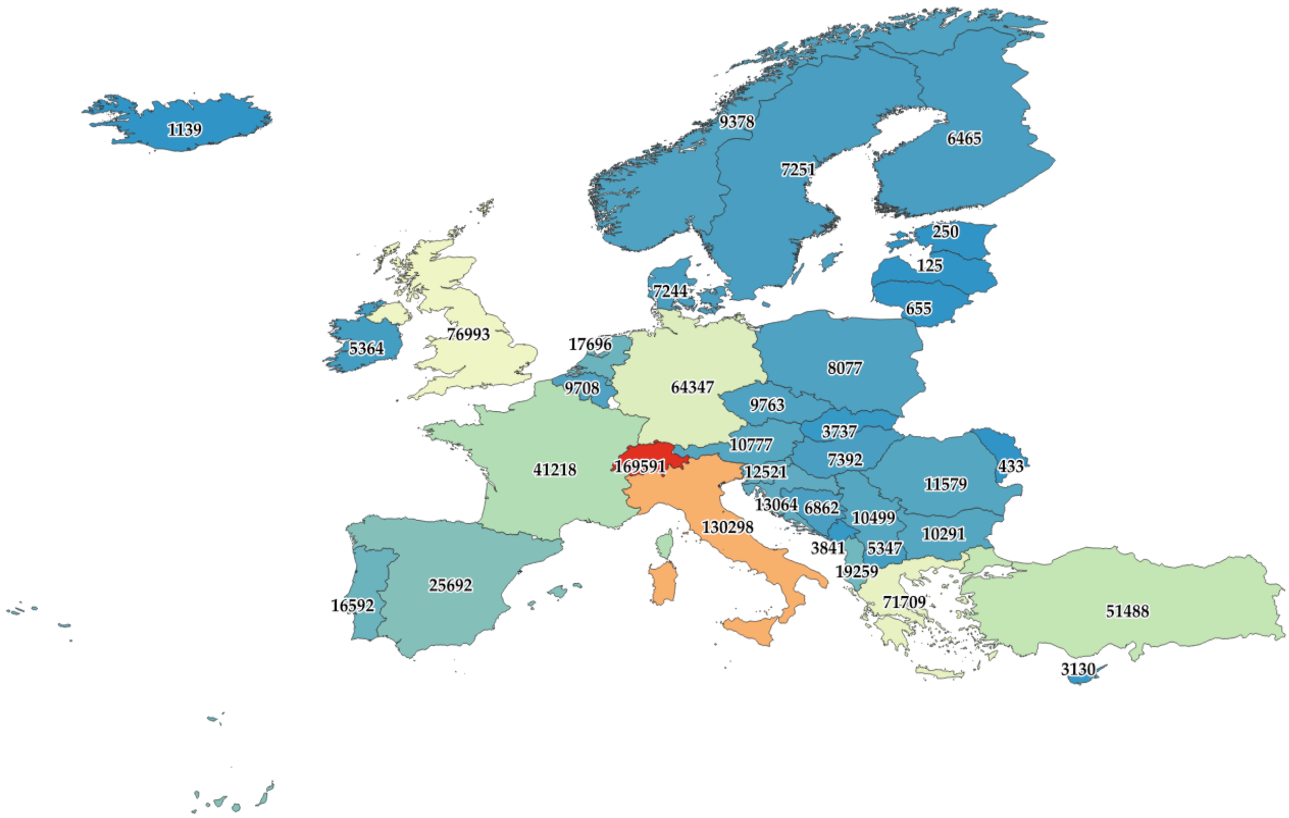
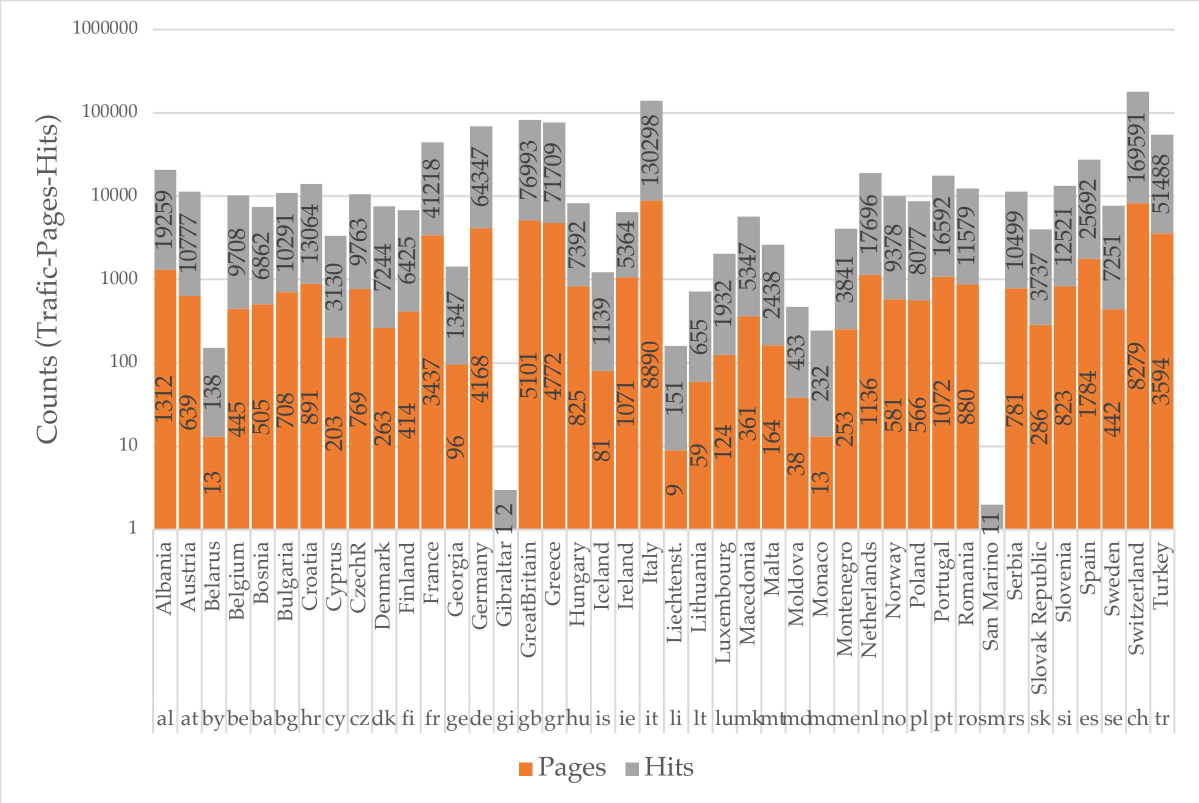


Figure 3: Spatial distribution of the user access of EFEHR web-portal per country in Europe between March 2019 to February 2020.


Counts (Trafic-Pages-Hits)
Pages
Hits

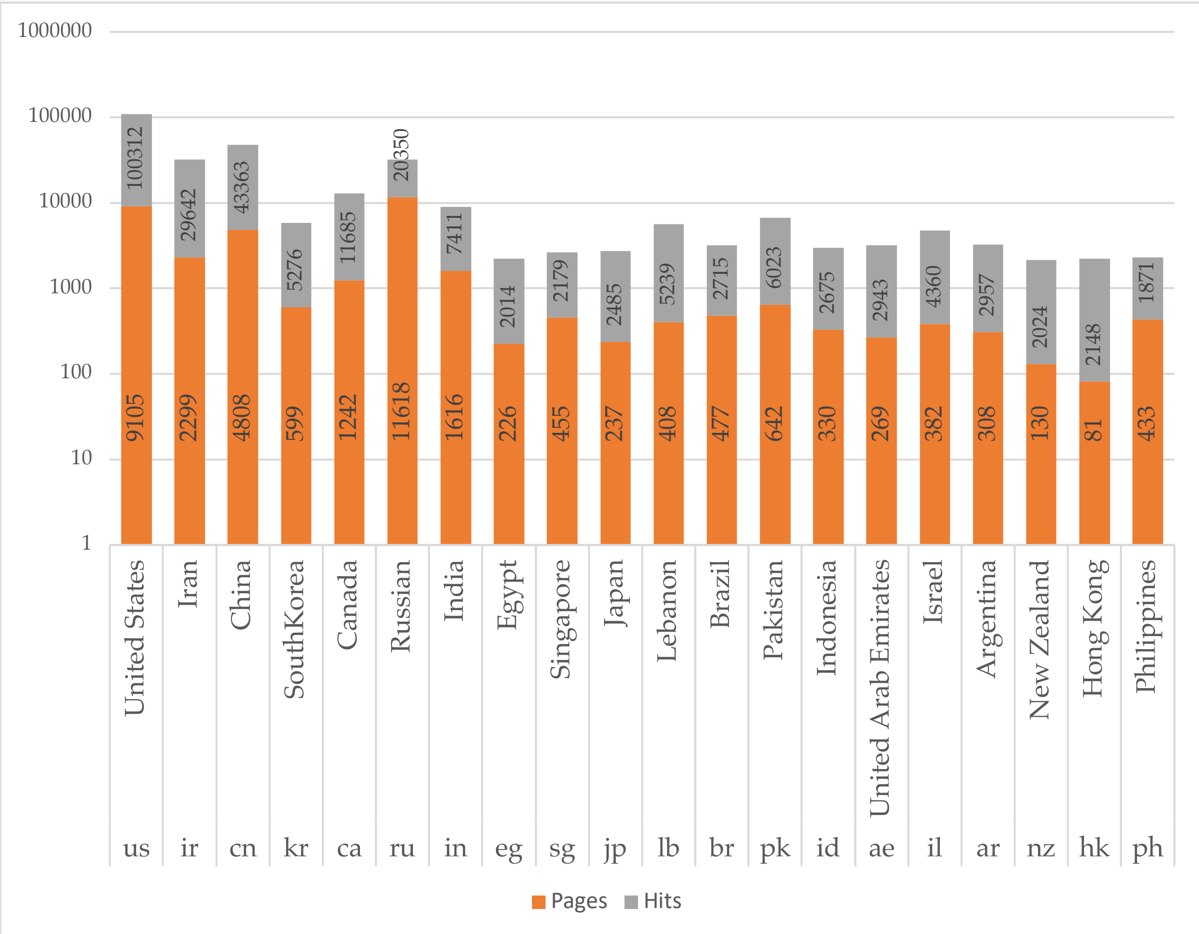

Pages
Hits

Figure 4: Frequency of accessing EFEHR web-portal for European Countries (top) and non-European Countries (bottom). The access stats combine jointly the access of the webpage and number of pages displayed in one-year period.

Country	Traffic (GB)	Pages	Page Hits
Switzerland	5.1	8279	169591
Italy	50.1	8890	130298
United Kingdom	15.04	5101	76993
Greece	12.67	4772	71709
Germany	15.98	4168	64347
Turkey	12.22	3594	51488
France	11.03	3437	41218
Spain	7.42	1784	25692
Albania	9.91	1312	19259
Netherlands	3.23	1136	17696
Portugal	7.93	1072	16592
Croatia	2.52	891	13064
Slovenia	1.53	823	12521
Romania	3.28	880	11579
Austria	2.45	639	10777
Serbia	3.56	781	10499
Bulgaria	1.44	708	10291
Czech Republic	4.59	769	9763
Belgium	2.81	445	9708
Norway	2.87	581	9378

Country	Traffic (GB)	Pages	Page Hits
United States	16.3	9105	100312
Iran	5.1	2299	29642
China	4.4	4808	43363
South Korea	2.1	599	5276
Canada	1.9	1242	11685
Russian	1.7	11618	20350
India	1.6	1616	7411
Egypt	1.2	226	2014
Singapore	1.0	455	2179
Japan	1.0	237	2485
Lebanon	0.6	408	5239
Brazil	0.5	477	2715
Pakistan	0.5	642	6023
Indonesia	0.5	330	2675
United Arab Emirates	0.5	269	2943
Israel	0.2	382	4360
Argentina	0.2	308	2957
New Zealand	0.2	130	2024
Hong Kong	0.1	81	2148
Philippines	0.4	433	1871

Table 1: Frequency of accessing EFEHR web-portal for European Countries (left) and non-European Countries (right). The access stats combine jointly the access of the webpage, number of pages and the generated web-traffic (in gigabytes) illustrated in Figure 3.

Data access (hazard results)

This section provides the information on how the web service API was used: operations to either retrieve hazard data, or meta-information (availability of data, covered parameter space etc.). Most data access refers to hazard maps (three times more interactions compared to both hazard curves and hazard spectra, Figure 5).

Source of the requests are interactive use of the viewer applications on the EFEHR portal, as well as direct, scripted service requests

The histogram plots in Figure 5 show access rates to the different web services per topic, from March 2019 to Spring 2020. These access rates are somehow constant over a year, however increased access can be observed in June 2019 for hazard maps and hazard spectra.

At this point, we are not able to explain the increased traffic but various factors can be considered: an occurrence of a significant earthquake in Europe, a user that repeatedly accessed and downloaded recursively hazard data, a specific project for a university class where the students have to access the online resources, traffic from the development of EPOS ICS services (https://epos-ip.org/data-services/ict-architecture/ics-tcs-communication) or access from interconnected web-platforms (OpenQuake https://platform.openquake.org/).

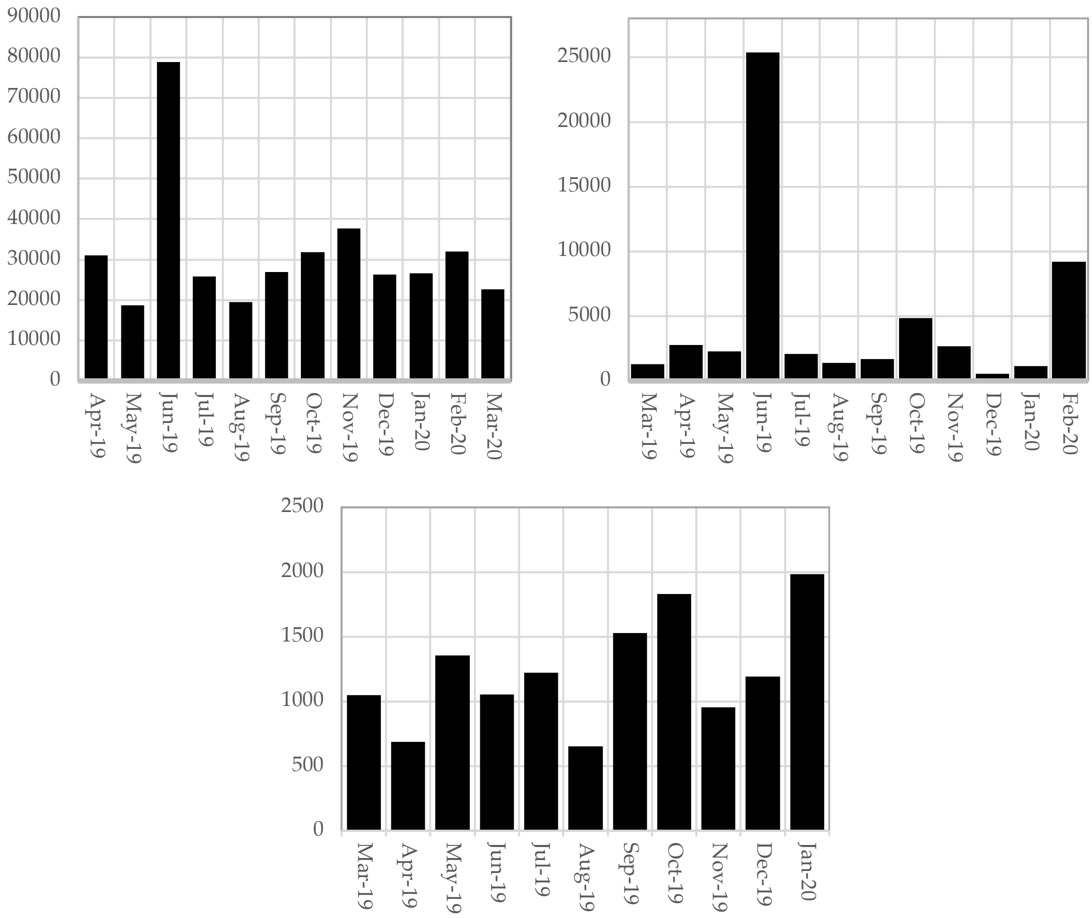


Figure 5: Histogram of the EFEHR hazard web-services. The most frequent web-portlet is the hazard map (top left), followed by the hazard spectra (top right) and hazard curves (bottom)

Impact Assessment

The data, models and results provided by the EFEHR platform have a strong impact on various communities from seismology to engineering as well as the public and decision makers. To quantify the impact of the EFEHR usage to the community, the relevant citations are queried from various scientific portals and listed below.

Note that the use of any dataset provided by EFEHR web-platform should be cited with the Giardini et al (2013) – online resources reference. However, there are many publications that refer to data obtained from EFEHR web-portal as data of SHARE Project (www.share-eu.org) without any reference or citation to neither Giardini et al (2013) or Wössner et al (2015).

With Google Scholar, the following numbers have been updated to reporting period March 2019 to March 2020:

90 citations of ESHM13 online resources available via EFEHR web-platform as Giardini et al (2013)
244 citations of the ESHM13 data sets as Wössner et al (2015)
130 citations of www.efehr.org
195 citations of data and models of www.share-eu.org

Other: EPOS/SERA Metadata and Data Policy (unchanged from D18.1)

All existing web-services of EFEHR are compliant with the EPOS DCAT-AP standard for machine discovery (https://github.com/epos-eu/EPOS-DCAT-AP) since 2019. For all webservices and data we are also working on adhering to the EPOS metadata format, ensuring that DOIs can be assigned to all products as well as CC BY SA v4.0 open data licenses, in order to meet the requirements of the EPOS Data Policy. The EFEHR Data Management Plan is same as in D18.1 and summarized in the next sections.

EFEHR Data Management Plan

<u>Data Summary</u>

EFEHR data and models are collected from completed scientific projects for long-term archiving, documentation, accessibility and use in research. Data covers the following domains:

Observed seismicity / earthquake catalogs;

Seismicity models, including parametrized active faults, seismic zones, and gridded seismicity models, ground motion prediction models, and associated logic trees;

Resulting probabilistic earthquake hazard data, expressed in hazard maps, curves and spectra;

Hazard model documentation;

Expert information and best practice documents for probabilistic hazard assessment.

All data comes from original research reports, but homogenized data formats and representations (Natural hazards' Risk Markup Language (NRML), OGC Web Map Services (WMS) response, geospatial vector shape file, open standard to represent seismological data (QuakeML, https://quake.ethz.ch/quakeml/), a variant of eXtensible Markup Language (xml) and tabular ASCII data). Documentation, products and services are mostly addressed to a well-informed public, including researchers and engineers.

<u>FAIR Data Policy</u>

Entire datasets identified by doi identifiers and full respective metadata sets. For access to individual data points, a set of EFEHR-specific discovery services (RESTful web service API) is offered. The API is documented in WADL and explanatory text at http://www.efehr.org/en/Documentation/web-services/. An EPOS-DCAT-compatible secondary documentation is in preparation and planned to be available by the end of 2018.

Data is accessible as bulk downloads - entire hazard models in OpenQuake input NRML format and maps (ESRI shape format), as individual data points, hazard curves and spectra discoverable and retrievable from a web service API in NRML format, and (for spatial data) via an OGC standard WMS (web map service) interface. Thus, all data is in well-documented community or industry-standard formats. Standard access services are used as far as available; otherwise standards have been defined for EFEHR. All data holdings are freely accessible to unregistered users under the license agreed on with the originating project / initial provider. EFEHR tries to homogenize agreements to the Creative Commons - CC BY SA v4.0 (https://creativecommons.org/licenses/by-sa/4.0) open data license.

Allocation of Resources

Standard technical operation, knowledge transfer, expertise and a basic infrastructure for quality-controlled seismic hazard/risk assessment are covered by EFEHR being one of the long-term strategic pillars of the European Plate Observatory EPOS, and backed up by ETH/SED. IT support and maintenance is covered by standard SED IT operations and the respective maintenance and 24/7 service team. Project specific scientific support, research task, and development of new services follow a long-term plan, but are financed on a project basis.

Data Security

Data preservation and disaster recovery is granted by two daily off-site backups of both data holdings and virtual service infrastructure, with a preservation time of 3 months. Database integrity, service stability and access control is granted by a 3-layer system architecture (database/data holdings <-> access services <-> web layer) with firewalled interconnects and full logging on the upper two layers.

Service continuity is supported by a 24/7 it monitoring and intervention team. However, no formal service availability level is guaranteed.

Ethical Aspects

EFEHR does not hold individual or personalized data in its scientific content, nor request or log such data from users.

References

Giardini G (1999), The global seismic hazard assessment program (GSHAP)-1992/1999, Annals of Geophysics 42 (6)

Giardini D. et al., (2013) Seismic Hazard Harmonization in Europe (SHARE): Online Data Resource, http://portal.share-eu.org:8080/jetspeed/portal/, doi: 10.12686/SED-00000001-SHARE, 2013.

Giardini D, DanciuL, Erdik M, Sesetyan K, Demircioglu MB, Akkar S, Gülen L, Zare M (2018)Seismic hazard map of the Middle East. Bulletin of Earthquake Engineering. 16(8),3567-3570, https://doi.org/10.1007/s10518-018-0347-3

Wiemer S, Danciu L, Edwards B, et al (2016) Seismic Hazard Model 2015 for Switzerland (SUIhaz2015), Official Report of the Swiss Seismological Service). doi: 10.12686/a2.

Appendices: Details of EFEHR Portal Design

The EFEHR web-portal consists of several components summarized hereinafter. The key components are:

- Database server – with Postgresql 9.3 (https://www.postgresql.org), postgis extensions and daily backup
- Java/Tomcat/OpenCMS (http://www.opencms.org/en/development/installation/server.html) based web content management system
- Standalone interactive data viewers in html/javascript (Ext Js, GeoExt, OpenLayers – see http://presentations.opengeo.org/2012_javascript/javascript/concepts.html)
- Map server (https://live.osgeo.org/archive/6.5/it/overview/mapserver_overview.html) implementing the OGC web map service standard.
- A REST (Representational State Transfer) Web Service provider implemented in Java (https://docs.oracle.com/javaee/6/tutorial/doc/gijqy.html)

DELIVERABLE

D20.1 Report on access statistics and service provision of VA1-VA5

(PART 5/IGPAS)

Work package	WP22 - VA5: Virtual Access to seismological products and information at IGPAS
Lead	IGPAS
Authors	Monika Sobiesiak, Stanisław Lasocki, Piotr Sałek, Dominika Wenc, Kostas Leptokaropoulos, Joanna Kocot
Reviewers	Marcelo Assumpção, University of São Paulo, R. Bossu (EMSC), K. Saleh (ETH)
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	29.02.2020
Submission date	30.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D20.1_Report_Virtual_Access_statistics.pdf

Table of Contents

Summary

The IS-EPOS platform (https://tcs.ah-epos.eu/) is a web-based utility which provides a unique collection of anthropogenic seismicity data, paired with industrial production data in hydrocarbon extraction, geothermal energy exploitation, underground mining, water reservoir impoundment and experimental data. The platform supports research through dissemination of the collected data to a wider community and through the development of specific software tools (applications) for statistical and waveform analysis. The platform structure offers an own web-based workspace for each user together with respective data handling, processing, resource management and visualization tools.

1 Introduction

The ever growing need of energy and natural resources for industrial production have stirred at the same time a growing awareness of its negative consequences: Damaging and disagreeable effects like induced or triggered seismicity in areas which are usually aseismic, water level changes, ground water contamination or even landslides (Froude[1] et al., 2018) are just some of the observed phenomena felt by populations or the wider public. As cause of these changes in Earths' structure, new exploitation methods like unconventional hydrocarbon exploitation through fracking methods, enhanced geothermal energy exploitation and CO2 sequestration, but also the "classical" underground and surface mining activities like conventional hydrocarbon extraction and the impoundment of surface reservoirs for liquids are widely recognized (Davies[2] et al., 2013, McGarr[3] et al., 2002, among others). In Figure 1, an overview on the published number of hazardous induced or triggered seismic episodes with a broad magnitude range between M1 and M7.6 assigned to the various anthropogenic activities are given. This histogram further shows, that induced or triggered micro- or large scale seismicity ranks among the most recognized and important responses to anthropogenic activities.

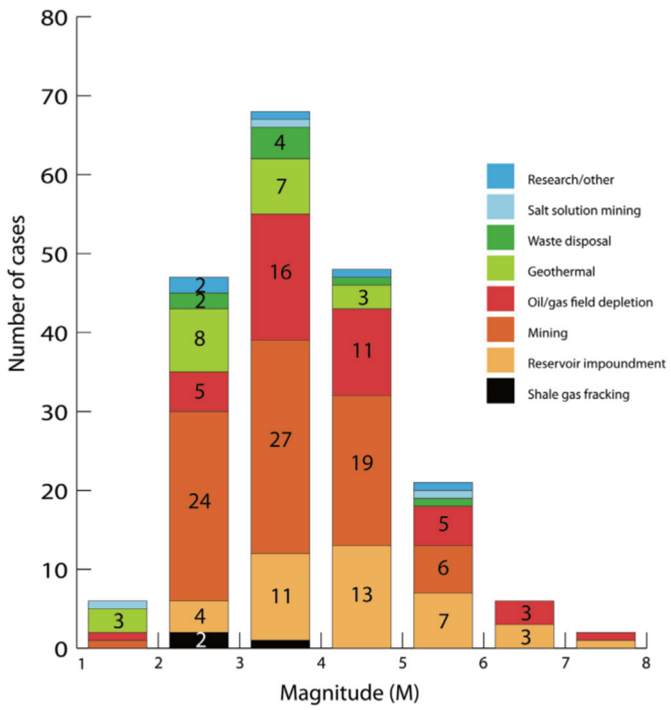


Figure 1: Published anthropogenic hazard cases since 1929, modified after Davies et al., 2013.

These increasing interactions between anthropogenic activities and concerned population requires more and more first-hand objective information about the observed response processes and open quality communication and exchange between involved groups such as scientists, engineers, industry, governmental and public entities, politicians and last but not least, the public.

Research in anthropogenic seismicity and hazard has been intensified in the last 30 years which led to new insights into the processes behind induced and triggered earthquakes associated with the above mentioned anthropogenic activities. New methods in monitoring, processing and analysis of data sets which take into account the specific conditions of monitoring underground, detection of micro-seismicity, analysis of source mechanisms of small earthquakes, often with low signal-to-noise ratios, could be brought on the way (Grigoli[4] et al., 2017, McGarr[3] et al., 2002). However, there are still pressing issues to be solved like improved hypocenter locations for small scale seismicity, real-time solutions for control systems in mines, discrimination of induced, triggered and tectonic seismicity to help on liability debates.

The IS-EPOS platform of Research into Anthropogenic Seismicity and other Anthropogenic Hazard (https://tcs.ah-epos.eu) is designed to serve the needs for research, information and knowledge transfer between science and industry, as well as expert information and education for the interested public. Within the above described versatile context, the platform is a unique collection of anthropogenic seismicity data sets combined with production data such as water injection rates, well head pressure and temperature. For investigations on causes for response processes, this data is indispensable. Such combined data sets are still sparse as industry data often is restricted because of private ownership or liability concerns. Therefore, dissemination of such kind of data through the platform to a wider community is a valuable contribution to the research in this field.

The platform, initially a product of IS-EPOS Polish national project, was further developed in the framework of EPOS IP H2020 as part of the Thematic Core Service of Anthropogenic Hazards (EPOS TCS AH, www.epos-ip.org). Within projects like SERA, the platform provides virtual open access to its resources, supporting in such a way research and investigation topics as targeted in SERA Joint Research Activities (JRA1-5).

1.1 Concept of the platform

As stated in the Introduction, the conceptual idea behind the platform is to serve as a utility which fosters interaction and exchange between the protagonists, science, industry, public and decision makers. This includes transfer of knowledge like scientific results i.e. from industry to science and vice versa, information on actual topics and events to the public and decision makers. In such a way, common projects could be enhanced where results contribute to the needs of all protagonists. This basic concept is sketched in Figure 2.

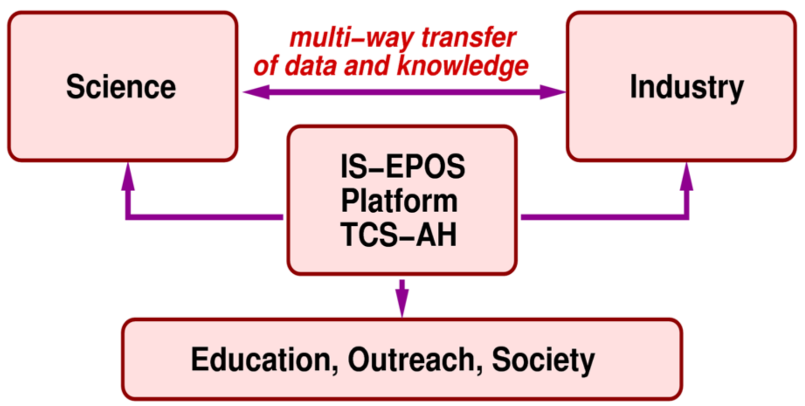


Figure 2: Sketch of the platform concept, afer Leptokaropoulos, 2017, EGU

1.2 Structure of the platform

Three major components form the structure of the platform: 1) the comprehensive data base of combined seismological, industrial and geologic data merged into episodes, where one episode describes the response to a single anthropogenic activity, 2) a pool of software and program packages to support research and analysis of respective data where each enclosed software contribution is called 'application', and 3) the web-based workspaces for each user to carry out respective studies and investigations with tools and data from the platform or own uploaded data. This structure is sketched in Figure 3. Furthermore, the platform hosts a comprehensive document repository of publications, manuals and overviews which are linked to the respective episodes and applications.

Physically, the platform is located at the Academic Computer Center CYFRONET at the AGH University of Science and Technology (ACC CYFRONET AGH) in Kraków, Poland, where it is technically developed and maintained through a team of IT scientists and engineers.

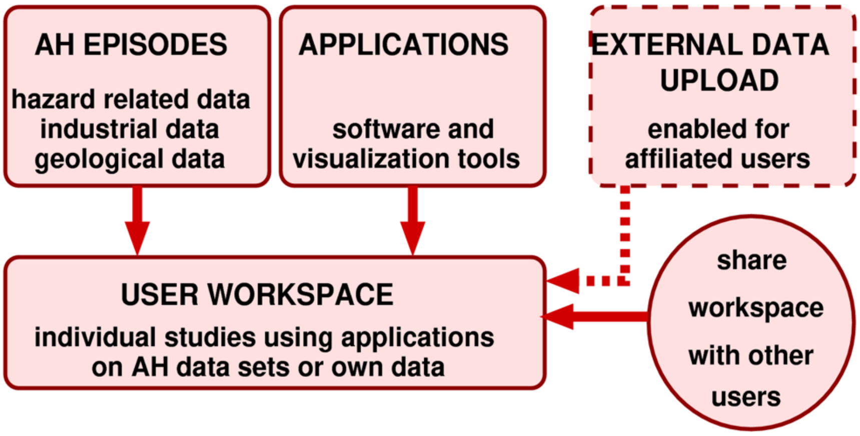


Figure 3: Major components of the platform structure after after Leptokaropoulos, 2017, EGU

1.3 Accessing the platform

The IS-EPOS web platform is available at https://tcs.ah-epos.eu/IS-EPOS, operational language is English. If a user has already an account, he/she can easily login by clicking on the **'LOGIN'** tab and simply providing the email address and password. To create an account, the user shall click on the **'SIGN-UP'** tab and complete the registration form. After submitting the registration form, the system verifies the provided affiliation and the user can hereinafter login the IS-EPOS web-platform. Currently, IS-EPOS platform offers two levels of access to the available material: first for anonymous users and second for registered users. An anonymous user can access information about available episodes, applications and browse the document repository. Anonymous users get access to data from the open episodes, but they will still have no access to the workspace nor data download. Registered users can access to all functionality available for anonymous users. Additionally, they get access to the workspace where they can use applications, share data and also download and upload data. An access to some episodes on the platform is restricted and available only to members of particular projects. In order to access such episodes, users need to request a project affiliation.

2 Components of the Platform

As mentioned before, the structure of the platform consists of three major components, which are 'Episodes' (anthropogenic hazard specific data sets, 'Applications' (software and tools for analysis and handling of data), and the 'Workspace' (for each user to combine data and applications to work on individual research targets).

2.1 Episodes

In total, there are 36 episodes containing seismological, industry production data and geological data (see Table 1). 24 of these episodes are open to all registered users and 12 episodes have still restricted access for members of respective projects. Each data set represents an episode related to an anthropogenic activity site. These sites are located in 12 different countries worldwide.

There are permanently new episodes from other sites and countries coming in, new data can always be added to the platform. However, certain quality tests are applied before the new data set is uploaded as well as formats are controlled or transferred when necessary to respective required formats. Within the reporting period of M19M34, the following 9 episodes have been made available:

Episode	*Type of activity*	*Country*
GISOS-Cerville	underground solution mining (exp.)	France
Pyhäsalmi Mine	in-situ underground laboratory	Finland
Starfish	underground gas storage	France
Wysin	shale gas exploitation	Poland
Monteynard	water reservoir	France
Cotton Valley	hydraulic fracturing	USA
Vouglans	water reservoir	France
Lacq Gas Field	conv. hydrocarbon extraction	France
Gazli	hydrocarbon field	Uzbekistan

Table 1: 9 new episodes which can be found on the platform. They are fully open to all users to whom virtual access is provided.

2.2 Applications

Programs, software and software packages are called 'Applications' on the platform, which are ready-to-use for data analysis. The user needs to transfer or upload a data set (whole episode or part of an episode) to their workspace and do the same with the application selected for individual studies. When the application is executed on the workspace, all results and resulting graphics will be displayed on the screen and stored as data files in the workspace. As an example, in Figure 4 (after Chapter 2), you can see a 3D distribution of seismicity in Bobrek Mine (Poland). Up to now, 57 applications are available for the users' data analyses studies. A list of new applications added to the platform in M19M34 is given in Table 2.

In general, the applications are sub-divided into four groups:

- data handling applications (include catalog filtering, extraction of parameters, seismic phase picking, conversion of formats, download and upload tools etc.)
- data processing applications (include autocorrelation tool, cross correlation, tool for focal mechanism and moment tensor calculations, inter-event time distribution analysis, estimation of maximum credible magnitude, etc.)
- resource management applications (include data catalog format change, i.e. from matlab to ascii, export of matlab files as XLSX spreadsheet, etc.)
- visualization applications (like histograms for mining front advances, 3D seismicity hypocenter distributions of which an example can be seen at the end of this chapter in Figure 4, seismicity related to mining front advance, etc.).

Application name
Anderson-Darling test for magnitude distribution
Test for multimodal magnitude distribution
Cluster analysis
Transformation to equivalent dimensions
MERGER-Dynamic risk analysis using a bow-tie approach
CSV to catalog converter
Catalog to ASCII converter
Basic vector operations
Earthquake interactions: Georesource scale
Ground Motion Prediction Equations: GMPE calculation
Ground Motion Prediction Equations: Final model analysis
Ground Motion Prediction Equations: : Residuals analysis
Template matching based detection algorithm
Earthquake interactions: Mainshock scale
Earthquake swarms: reshuffling analysis
Time correlated earthquakes (seasonal trends)
Mechanisms: Full moment tensor
Mechanisms: Shear slip
Mechanisms: Shear-tensile cracks
Coda wave interferometry detection of velocity changes
Waveform based seismic event location
Estimate of maximum possible magnitude for reservoir triggered seismicity

Table 2: List of new applications added in M19M34

Most of the software codes behind the applications are open source which is the general policy for software contributions on the platform. A number of researchers from different international institutions have contributed to the application list.

2.3 User's Workspace

The *workspace* is the user's individual area to work on with the data sets with software tools provided by the platform. However, the user also has the option to download data to an own device or upload additional data to the workspace from an external account. Thus, the workspace provides to the user a web-based tool for data analysis which can be used interactively on episode data from the platform or on own data. In order to start an analysis, both, the selected data plus the chosen application have to be transferred actively to the workspace. Here, software tools and data sets can be stored in different directories which are defined by the user. The applications are activated in the workspace, where results are displayed as single values in files or as graphs shown after a successful run. Both output options can be either stored in the workspace directories or can be downloaded to own devices. Summarizing the advantages, the workspace provides:

- fast overviews on data properties and parameters
- upload and download tools
- storage for data and results
- flexible usage of applications
- detailed analysis of data
- visualization of data and results.

In order to facilitate collaboration for projects or also individual co-operations between researchers, a share function has been implemented. With this share function, users can provide access to selected parts of their own resources to other researchers on the platform. This access can be bi-directional.

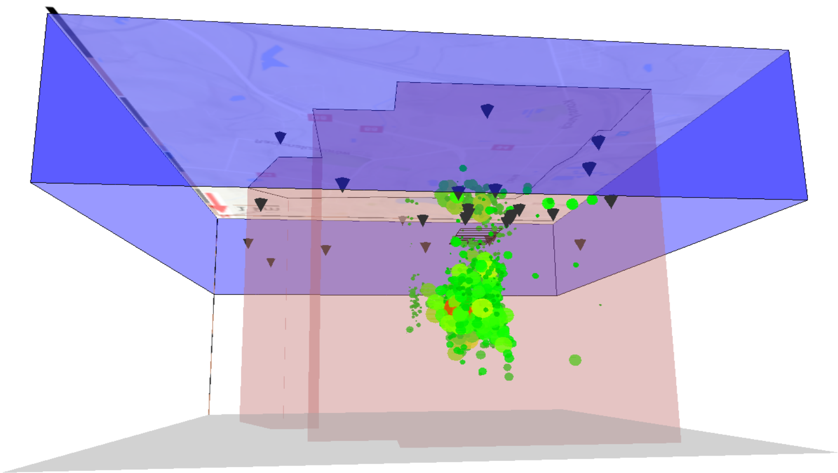


Figure 4: Example of clustered induced seismicity from Bobrek Mine (Poland). The light red colour marks the outer limits of the mine; the light purple rectangle gives one layer of the velocity model; green, yellow, and red circles mark the earthquake hypocenters. The figure was generated with the help of the 3D visualization tools of the platform.

3 Summary of VA5 Statistics

By the start of the SERA project, IG PAS and ACC Cyfronet AGH ensured a virtual access to the IS-EPOS platform. Until that time, the number of registered users was in total 564. Before and also within the reporting period of M19 to M34, statistics show that the number of users, the number of registrations, the number of episodes and also the number of applications are constantly increasing.

3.1 User statistics

The following histograms and the maps of the global distribution user registrations and logins demonstrate the continuous increase in all usages and services.

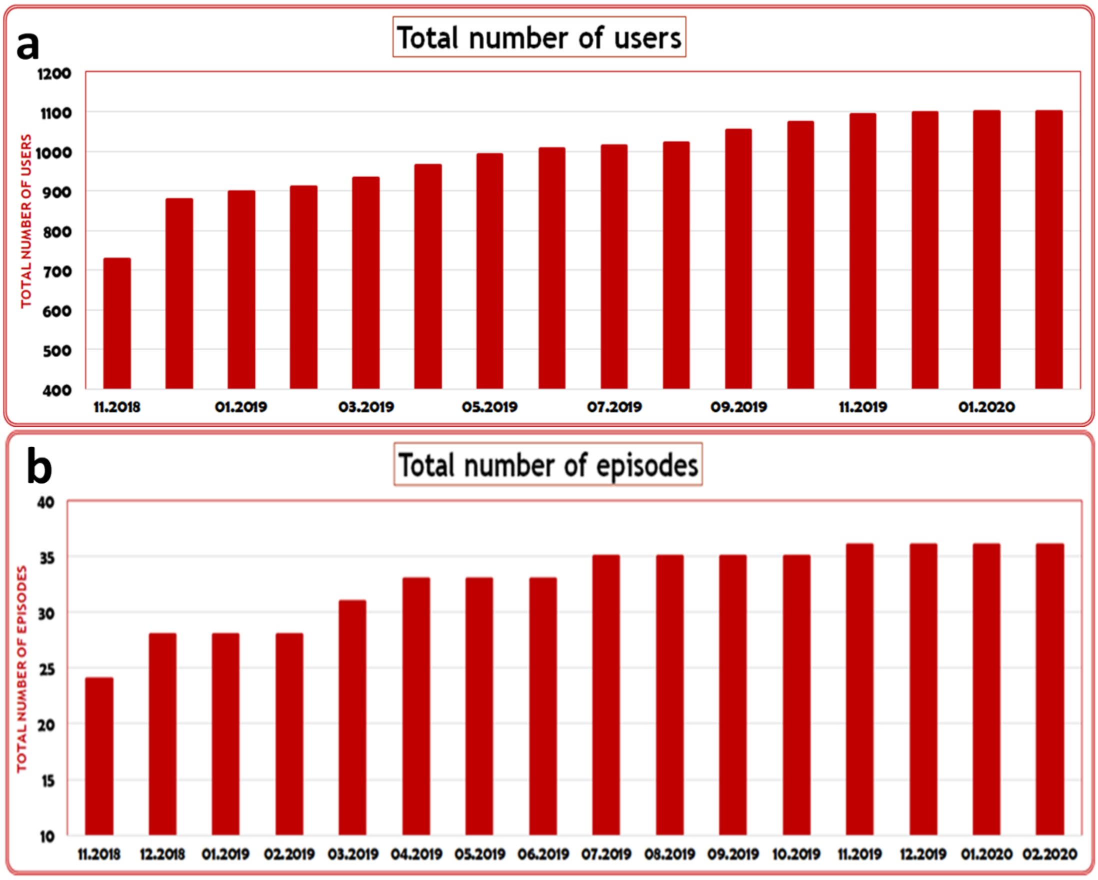


Figure 5 a) shows that the maximum number of users on the platform is 1100. Also the numbers of contributed episodes and data sets have been rising (**Figure 5b**) which might be linked to the growing internationality of the users. This is demonstrated by the map of registrations where users from Asia, Southern and Northern America, Africa and Australia have registered on the platform (see **Figure 8**).

In the following **Figure 6**, the frequency of use of individual applications shows that 'CatalogFilter' and 'SignalDownload' are the two favourite applications for the users, followed by 'Trmloc' (software for locating events) and 'SpectralAnalysis' tools. While the succession of frequency of use has rarely changed, the maximum use of each application has increased in comparison to the former estimation in the report M1M18.

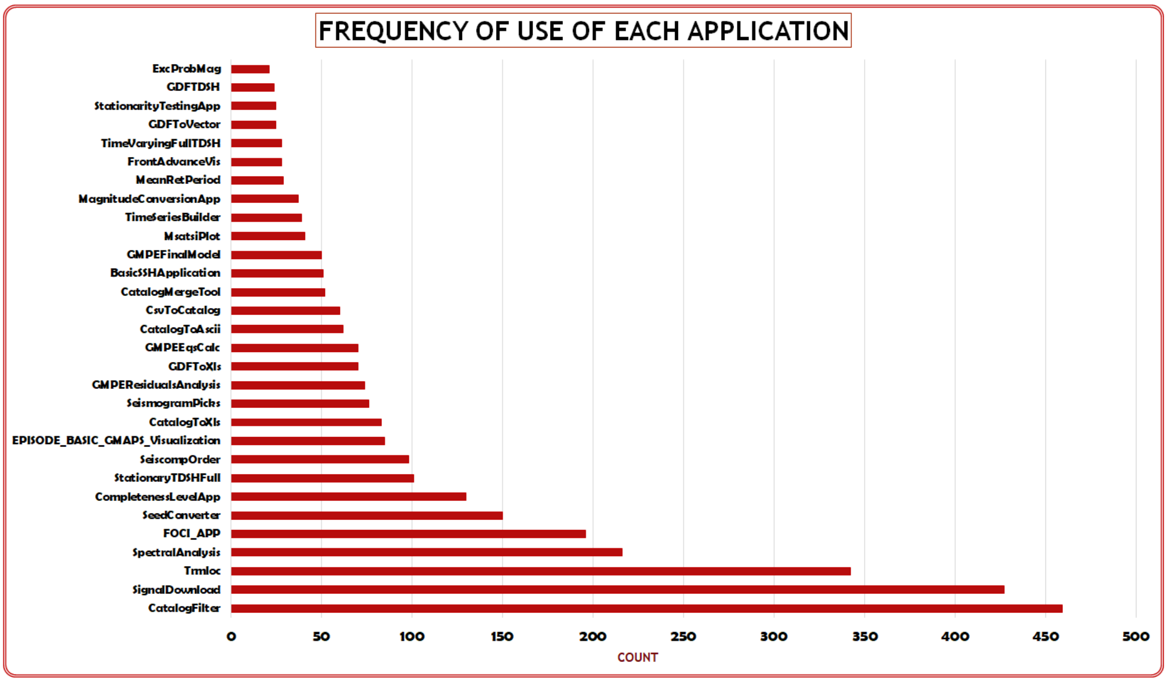


Figure 6. Frequency of Application Use.

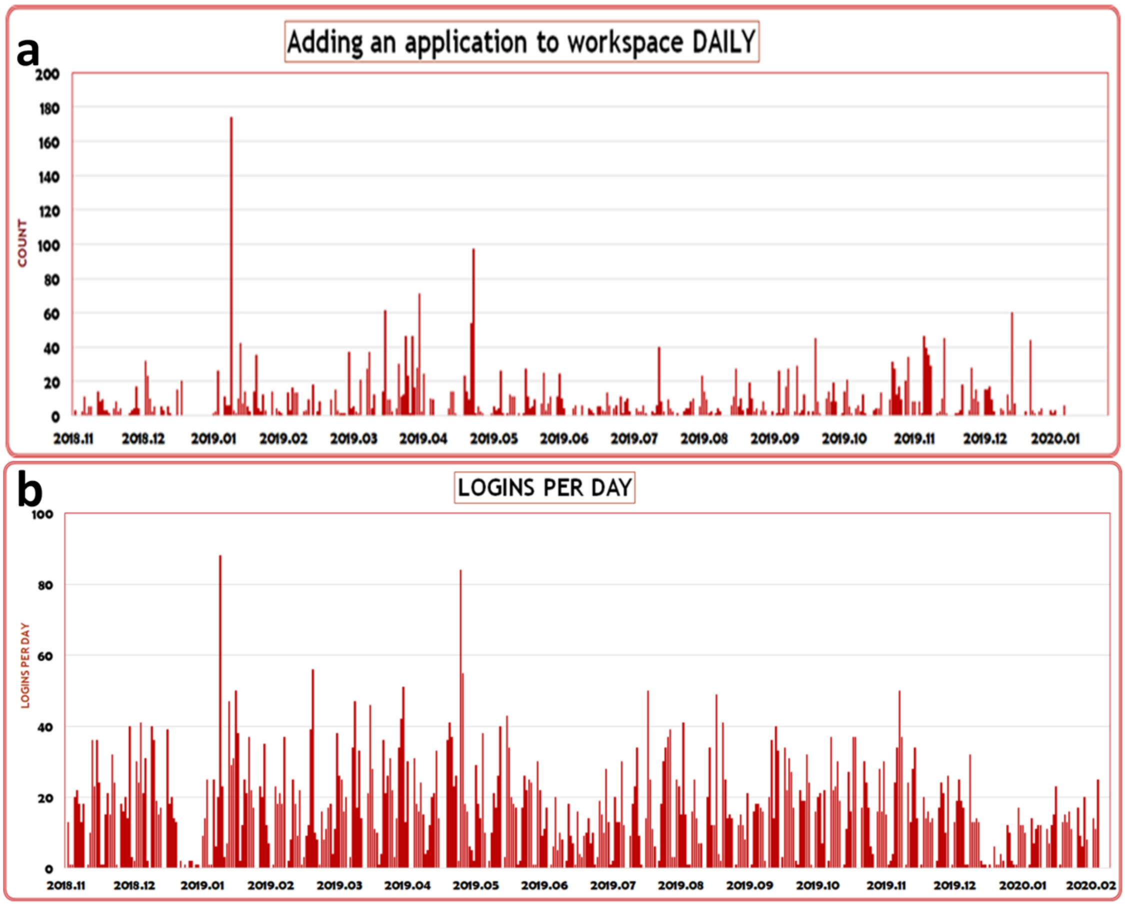


Figure 7 a) and b): While previous statistics show a permanent increase in services usage, Figure 7a) and b) show a more complex behaviour in daily usage of applications and general logins.

3.2 Concluding remarks and outlook

The statistics as shown in the former Chapter point out that the IS-EPOS Platform of Induced Seismicity and Anthropogenic Hazard has become a considered tool in the area of anthropogenic seismicity and hazard cases. All important statistical values experienced a significant increase during the last 12 months. To summarize the advantages of the platform use, the following listing is given. The platform provides:

- an overview on anthropogenic hazard cases with induced or triggered seismicity involved
- tools for downloading open data and uploading own data to the respective workspaces
- options to change formats both for waveform files as well as matlab result files
- open source applications
- share function for sharing data or actual results with other researchers on the platforms
- visualization tools for 3D animation to achieve information on the spatial and temporal distribution of seismicity.

With the aim to further enhance the flexibility of the platform for advanced requirements of the users, the following issues will be targeted in the future:

- providing a tool for creating shared workspaces among members of specific projects
- facilitating interactive work on developing new applications
- possibility for downloading applications for stand-alone use on own computers.

Following the recommendation of our reviewer, we would like to take the challenge to envisage future comparative studies from sites worldwide. Aiming at these studies, like comparing seismicity or production parameters from mines in Chile to similar mining activity in Europe, might help to substantiate general findings with universal character for describing anthropogenic processes. One effort in this direction is the internationalization of the platform with worldwide users and episodes. The following Figure 8 shows the first attempts in this direction.

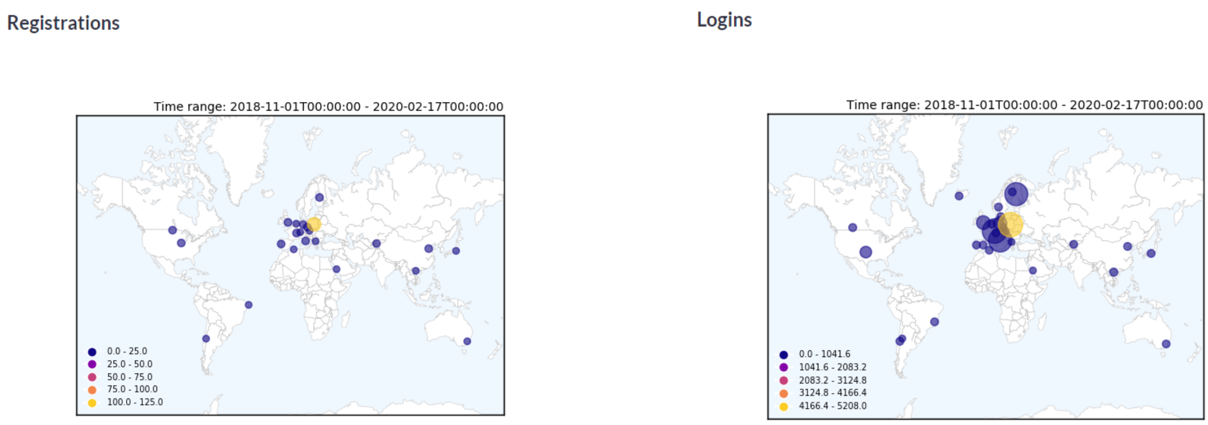


Figure 8: Maps of global user registrations and logins.

4 References

1) Froude, M.J., and Petley, D.N., Global fatal landslide occurrence from 2004- 2016, Nat. Hazards Earth Syst. Sci.,18, 2161-2181, https://doi.org/10.5194/nhess-18-2161-2018, 2018.

2) Davies, R., G. Foulger, A. Bindley, P. Styles, Induced seismicity and hydraulic fracture for the recovery of hydrocarbons, Marine and Petroleum Geology, 45, 171-185, http://dx.doi.org/10.1016/j.marpetgeo.2013.03.016, 2013.

3) McGarr, A., D. Simpson, and L. Seeber, Case histories of induced and triggered seismicity, Int. Geophys. Ser., 81(A), 647-664, 2002.

4) Grigoli, F., S. Cesca, E. Priolo, A.P. Rinaldi, J.F. Clinton, T.A. Stabile, B. Dost, M.G. Fernandez, S. Wiemer, and T. Dahm, Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective, Rev. Geophys., 55, doi:10.1002/2016RG000542, 2017.

5) Leptokaropoulos, K., and the IS-EPOS Team, EPOS Thematic Core Service (TCS) – Anthropogenic Hazards (AH), Demonstration Session of the IS-EPOS Platform in EGU General Assembly, 2017.

Contact

Project lead	ETH Zürich
Project coordinator	Prof. Dr. Domenico Giardini
Project manager	Dr. Kauzar Saleh
Project office	ETH Department of Earth Sciences
	Sonneggstrasse 5, NO H-floor, CH-8092 Zürich
	sera_office@erdw.ethz.ch
	+41 44 632 9690
Project website	www.sera-eu.org