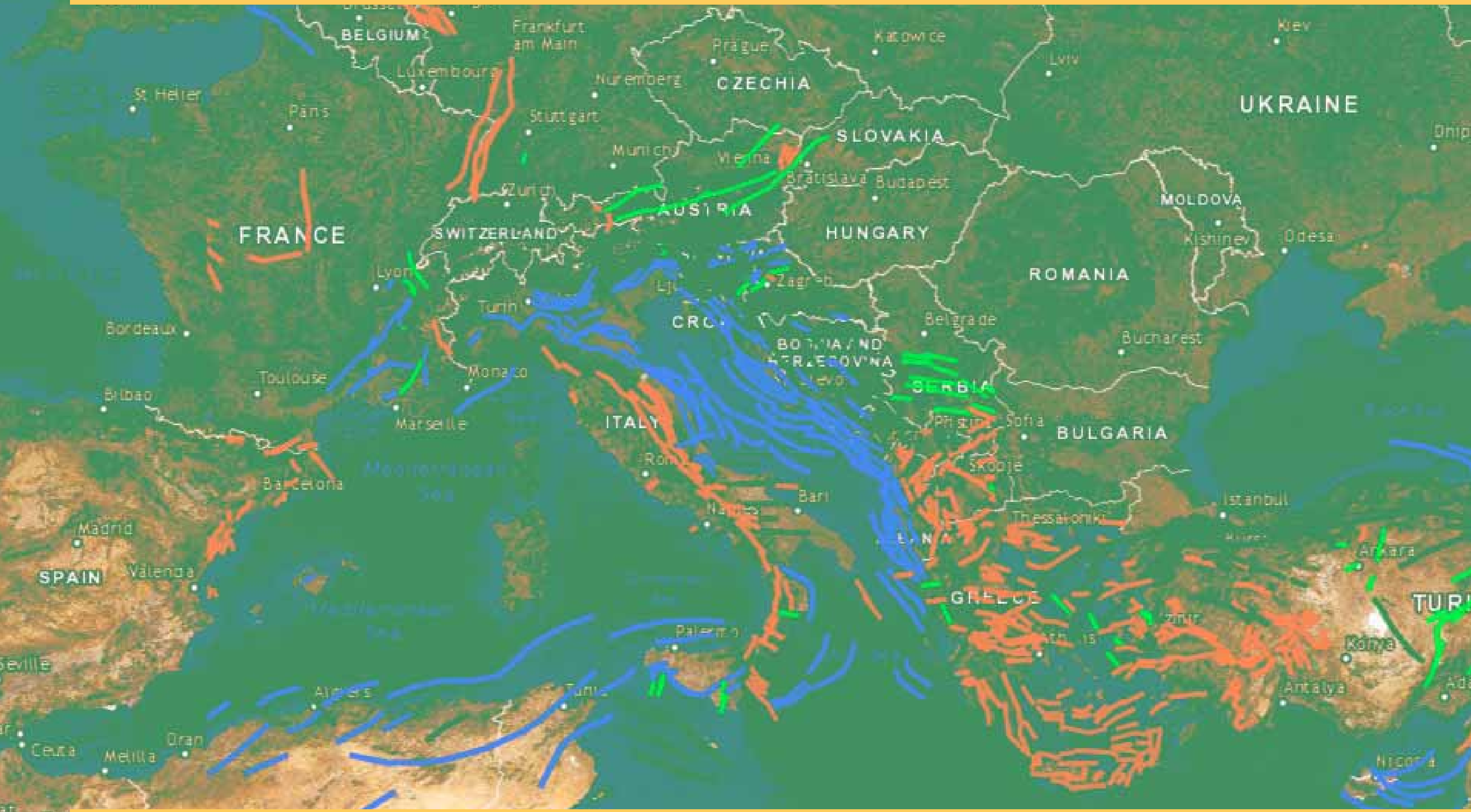
ISSN 2039-7941
anno 2023 numero 474



RAPPORTI TECNICI INGV

Technical documentation of SEISMOFAULTS.EU: the IT infrastructure employed by the European Databases of Seismogenic Faults (EDSF) installation

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

474

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

Sede: Via di Vigna Murata, 605 | Roma

ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

ISSN 2039-7941
anno **2023** numero **474**

RAPPORTI TECNICI INGV

Technical documentation of SEISMOFAULTS.EU: the IT infrastructure employed by the European Databases of Seismogenic Faults (EDSF) installation

Roberto Vallone*, Roberto Basili

INGV | Istituto Nazionale di Geofisica e Vulcanologia, Sezione Roma 1

**Corresponding author*

Accepted 15 September 2023 | *Accettato 15 settembre 2023*

How to cite | *Come citare* R. Vallone, R. Basili, (2023). Technical documentation of SEISMOFAULTS.EU: the IT infrastructure employed by the European Databases of Seismogenic Faults (EDSF) installation. Rapp. Tec. INGV, 474: 1-34, https://doi.org/10.13127/rpt/474

Cover | *In copertina* The European Fault-Source Model 2020; image processing by B. Angioni

474

INDEX

Abstract

SEISMOFAULTS.EU is the IT infrastructure designed and implemented to publish datasets that are part of the European Databases of Seismogenic Faults (EDSF) installation hosted by the Istituto Nazionale di Geofisica e Vulcanologia.
It consists of carefully selected and configured hardware and software in order to ensure a reliable, and secure service. The implemented backup solutions and continuous monitoring of the entire platform at various levels safeguard the system from disruptions due to various types of possible malfunctions.
The main functionalities of SEISMOFAULTS.EU include the publication of datasets of seismogenic sources of the European and Mediterranean area through the standard web services developed by the Open Geospatial Consortium. Websites related to the same datasets as well as web services related to sibling projects, are also published through the infrastructure.
Following government recommendations for public administration, most of the software used in the platform has an open-source license. The main software applications involved in web services publishing are PostgreSQL/PostGIS and GeoServer.
SEISMOFAULTS.EU was designed under the auspices of EPOS, Thematic Core Service Seismology, and actively contributed to it by integrating datasets on seismogenic faults in the Euro-Mediterranean area into the EPOS Integrated Core Service - Central hub platform.
The combination of efforts between EPOS and SEISMOFAULTS.EU strengthens the scientific geoscience community in coordinating efforts aimed at advancing knowledge in the field of seismology and the study of seismogenic sources and earthquake hazard analyses.

Riassunto

SEISMOFAULTS.EU è l'infrastruttura informatica progettata e implementata per pubblicare i dataset che fanno parte dell'installazione European Databases of Seismogenic Faults (EDSF) ospitata dall'Istituto Nazionale di Geofisica e Vulcanologia.
Si tratta di hardware e software accuratamente selezionati e configurati per garantire un servizio affidabile e sicuro. Le soluzioni di backup implementate e il monitoraggio continuo dell'intera piattaforma a vari livelli salvaguardano il sistema da interruzioni dovute a vari tipi di possibili malfunzionamenti.
Le principali funzionalità di SEISMOFAULTS.EU comprendono la pubblicazione di dataset di sorgenti sismogenetiche dell'area europea e mediterranea attraverso i servizi web standard sviluppati dall'Open Geospatial Consortium. Anche i siti web relativi agli stessi set di dati e i servizi web relativi ai progetti fratelli sono pubblicati attraverso l'infrastruttura.
Seguendo le raccomandazioni governative per la pubblica amministrazione, la maggior parte del software utilizzato nella piattaforma ha una licenza open-source. Le principali applicazioni software adottate per la pubblicazione di servizi web sono PostgreSQL/PostGIS e GeoServer.
SEISMOFAULTS.EU è stato progettato sotto gli auspici di EPOS, Thematic Core Service Seismology, e vi ha contribuito attivamente integrando i dataset sulle faglie sismogenetiche dell'area euro-mediterranea nella piattaforma EPOS Integrated Core Service - Central hub.
L'unione degli sforzi tra EPOS e SEISMOFAULTS.EU rafforza la comunità scientifica delle geoscienze nel coordinare gli sforzi volti a far progredire le conoscenze nel campo della sismologia e dello studio delle sorgenti sismogenetiche e delle analisi della pericolosità dei terremoti.

Keywords EPOS; Seismogenic faults; Web services | *Faglie sismogenetiche; Servizi web*

Introduction

The European Databases of Seismogenic Faults (EDSF) installation operates under the auspices of the EPOS TCS-Seismology [Haslinger et al., 2022] work program, particularly those of the European Facilities for Earthquake Hazard and Risk (EFEHR Consortium; http://www.efehr.org), and considers the principles expressed by the EPOS and INGV Data Policies [Anonymous, 2018; Puglisi et al., 2016]. The EDSF installation offers services that distribute datasets about seismogenic faulting proposed by the scientific community or solicited to the scientific community or stemming from project partnerships that involved the use or development of the EDSF installation itself. EDSF also hosts a large part of the data behind those datasets. Such datasets are published through the open standards developed by the Open Geospatial Consortium (OGC). The EDSF installation front end is a web portal (https://seismofaults.eu).
This document presents the IT infrastructure behind EDSF, named SEISMOFAULTS.EU, which is hosted by the Istituto Nazionale di Geofisica e Vulcanologia (INGV).
The scope of the SEISMOFAULTS.EU infrastructure is to provide IT support to the collection, storage, and publication, through web services, of EDSF datasets but also hosts a few other portals and websites to provide access to datasets related to sibling projects, such as the EPOS candidate TCS Tsunami [Babeyko et al., 2022], distributed through the *tsunamidata.org* portal. For example, SEISMOFAULTS.EU distributes the tsunami hazard model NEAMTHM18 [Basili et al., 2018; Basili et al., 2021] or the web services of the INGV *Catalogo dei Forti Terremoti in Italia (CFTI)* [Guidoboni et al., 2018; 2019]. The infrastructure includes various hardware and software components that work together to ensure the efficient and reliable operability of distributed databases and web portals. The hardware components include servers, storage devices, and networking equipment, which are housed in two separate INGV data centers, one in Rome and the other in Bologna. The servers are configured to run various software applications, including database management systems, web servers, and multiple data processing and analysis tools. The storage devices are used to preserve the databases, backups, and associated files, while the networking equipment connects the various hardware and software components.
The software components, in addition to the operating systems, include various open-source applications, data processing, and analysis tools. These applications are configured and customized to meet specific requirements, including data validation, storage, and visualization.
The infrastructure also includes several security measures, such as a firewall, intrusion detection systems, and access controls, to protect data and prevent unauthorized access or data breaches.
Overall, SEISMOFAULTS.EU is an articulated IT infrastructure designed to be scalable, flexible, and robust. It is also continuously monitored and maintained to ensure its reliability and availability.

1. Design and Implementation

1.1 Infrastructure design

The SEISMOFAULTS.EU architecture was initially designed in early 2016, first implemented in the second half of 2016, and has gone through several phases of update since then. At first, a single virtual machine (VM) was configured to cover the minimum publishing needs of some services related to datasets of national coverage. Starting in 2017, with the full involvement in the activities of the EU H2020 project EPOS - Implementation Phase, new hardware resources were gradually acquired to comply with the new requirements, and the architecture became increasingly complex until it reached the state that is described in this report.
Considering the primary SEISMOFAULTS.EU scopes, we set the reliability and simplicity of management as the two main goals to be achieved, which are also mutually related.

In an earlier configuration, we separated tasks according to a criterion based on task-specialized VMs (e.g., Database Management, Service Publishing, Website Publishing). This configuration was abandoned because whenever an upgrade was needed or a problem occurred on a VM (see Section 1.3), all published services became unavailable until the necessary tasks were completed. In the current configuration (Figure 1), everything that is published is structured according to the microservices paradigm (see Section 1.4). VMs are still present but are distinguished only by general criteria (e.g., a machine is dedicated to publishing seismogenic fault datasets, another machine is dedicated to websites, another machine is dedicated to publishing datasets related to sibling projects, and so on).

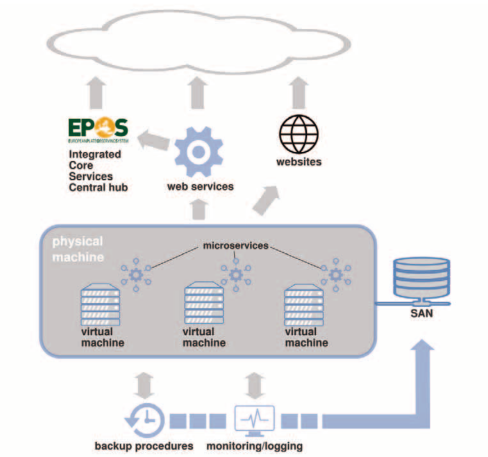


Figure 1 General scheme illustrating the main components of the SEISMOFAULTS.EU infrastructure configuration. The individual components are described in dedicated Sections of the main text.

1.2 Hardware

SEISMOFAULTS.EU is composed of three dedicated physical servers:

- N. 1 Hewlett Packard Enterprise (HPE) ProLiant BL460c (Gen9) equipped with two Intel® Xeon® CPU E5-2640 v4 @ 2.40GHz processors with 10/10 cores, 20 threads; 128 GB of RAM for every CPU with a frequency of 2133 MHz; two HDD HPE of 300GB in Raid 1. This server is entirely dedicated to development and testing.
- N. 2 HP ProLiant DL560 (Gen10) equipped with 4 Intel® Xeon® Gold 5118 CPU @ 2.30GHz 12/12 cores, 24 threads processors with 64 GB of RAM for every CPU with a frequency of 2400 MHz; four HDD da 300 GB in Raid 5 and 4 SSD da 1920 GB in RAID 5. These two servers host the production side of the infrastructure.

Besides the storage installed on board the server, additional storage space is provided by a Sinology DS718+ Network Attached Storage (NAS) with 2TB space (RAID 1) dedicated to single-user and working group backups and 50 TB of storage shared by the Storage Area Network (SAN), through the NFS protocol.
The ProLiant BL460c, one of the two DL560, the NAS, and the SAN are hosted in the Centro Servizi Informativi (CSI), i.e., the data center in the INGV premises in Rome, whereas the other DL560 is hosted in the INGV premises in Bologna.
Placing the production servers in two different geographic locations gives us the ability to keep our web services reachable not only in case of malfunction of our main server but also in case of extraordinary maintenance or failure of the entire IT infrastructure in the Rome premises.
Maintaining synchronization between the data published by the two nodes is carried out manually through a VPN connection. A choice was made not to implement an automatic synchronization system since all datasets published are currently static datasets, and the

addition of updated versions (which do not replace previous versions) occurs at monthly-to-yearly intervals.
The CSI provides and maintains the power supply and network connectivity.

1.3 Software

Following the governmental guidelines for software acquisition and reuse for the Italian Public Administrations [AGID, 2019], SEISMOFAULTS.EU implements almost exclusively open-source software.
In each physical server, we installed the Xen hypervisor for creating and managing virtual machines (Figure 2; Table 1).

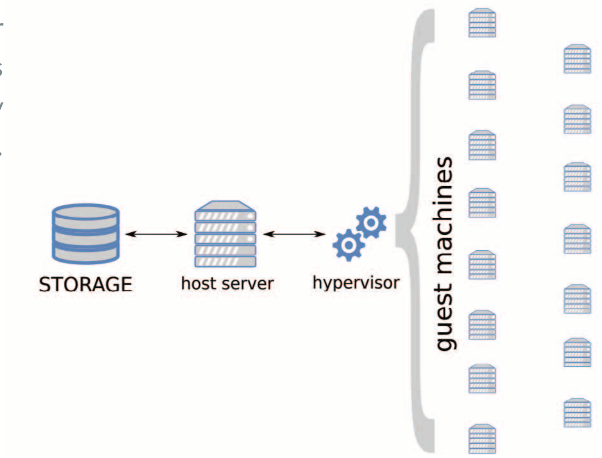


Figure 2 Every physical server is equipped with a hypervisor and linked with external storage. The hypervisor manages several virtual machines (guest machines), transparently distributing hardware resources.

A hypervisor is a software that enables the virtualization of computer resources. It allows multiple operating systems to run simultaneously on a single physical machine by abstracting and managing the underlying hardware, creating virtual machines (VM) that can share the same physical resources. The hypervisor provides isolation and resource allocation to ensure efficient and secure utilization of the host system.
In this way, each hypervisor-provided physical machine hosts several VMs with different distributions of the Linux OS, primarily Debian Linux, devoted to activities related to the frontend and backend of data distribution, and a VM with Microsoft Windows Server 2016 devoted to running applications, such as desktop GIS, legacy software, and technical tools, related to specific backend data manipulation, harmonization, and dataset organization (Figure 3).
A complete list of the software installed on SEISMOFAULTS.EU is provided in Table 1 while a description of the single configurations is provided in the following sections.

Software	Description	Website	Impl. level
Apache 2.0	Web server	https://apache.org	D
Apache Tomcat	Web Application Server	https://tomcat.apache.org	D
AWStats	Log file analyzer	https://awstats.sourceforge.io	D
Debian Operating System	Linux based operating system	https://debian.org	B
Docker	Containerization platform	https://docker.com	O

Fail2ban	Dynamic firewall rules adjustment	https://www.fail2ban.org	O
GeoServer	Geospatial data management and publication software	https://geoserver.og	D
Google Analytics	Web analytics service by Google LLC.	https://analytics.google.com	N/A
Joomla!	Web Content Management System	https://joomla.org	D
MariaDB	Relational DataBase Management System	https://mariadb.org	D
Microsoft Windows Server 2016	Operating System	https://microsoft.com/	O
Nagios Core	Monitoring software	https://nagios.com	D
NGINX	Web Server	https://nginx.com	D
PostGIS	Geographic extension library for PostgreSQL	https://postgis.net	D
PostgreSQL	Relational DataBase Management System	https://postgresql.org	D
Ubuntu Linux	Linux based operating system	https://ubuntu.com	B
WordPress	Web Content Management System	https://wordpress.org	D
Xen	Hypervisor	https://xenproject.org	N/A

Table 1 Software used by SEISMOFAULTS.EU installation. The 4th column indicates whether the software is used at OS level (O), dockerized (D), both (B), or not applicable (N/A).

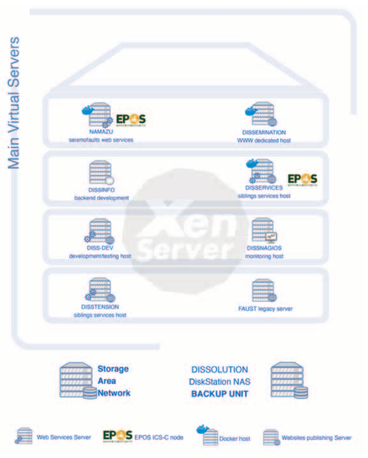


Figure 3 SEISMOFAULTS.EU main virtual machines and their primary tasks.

1.4 Docker containers

To enhance efficiency, start-up speed, and security, all services and websites are containerized through Docker containers technology.

Docker is a virtualization technology that creates isolated applications or services in an autonomous environment named "containers". This environment includes all the applications/services needed to be executed: dependencies, libraries, and configuration files.

Despite some similarities, containers differ from VM for several reasons:

1. They share the same kernel and operating system, making them more efficient and less resource-intensive than VMs.

2. They can be created, deleted, and ran faster, making them ideal for implementing microservices and applications that need few resources and great scalability.
3. They are portable. Once created and customized, they can be deployed and ran on almost every platform that can run the Docker runtime engine.
4. They can be easily deployed in the cloud or mixed environments.
5. They offer higher security. A carefully configured container is an environment isolated from the host OS and other containers. A malfunctioning or compromised container does not affect the functionalities of the other containers or the host OS.

Scalability is achieved as needed through a manual procedure to create additional containers (horizontal scalability) in case of excessive and prolonged load (reported through the monitoring system described in Section 2.3). This choice is based on the utilization statistics (Section 2.3) in recent years that have rarely shown appreciable slowdowns, even during the highest load periods. The justification is to be found both in the general efficiency of the system, whose configuration has been refined over the years, and in the fact that the published datasets are, in general, not particularly onerous in terms of data volume and, consequently, load on the system when querying services. In fact, the data volume ranges from a few Mb for DISS to a few tens of Mb for NEAMTHM18.
The typical file/folder structure of a "dockerized" service in SEISMOFAULTS.EU can be summarized as follows:

- A leading directory "to hold them all" with a name reminiscent of the name of the service. Often this directory is also a GIT repository (see below).
- A *docker-compose.yml* file (commonly known as *compose file*) containing details about the applications, networks, and volumes for setting up the application's environment. This file is used to create, deploy, and manage service-related containers.
- A directory containing *named volumes*. Volumes store and manage data generated and used by Docker containers. Volumes exist outside the Docker container and can preserve data even if the container is destroyed or moved. They typically store and share data between and among multiple containers or with the host OS (Figure 4).

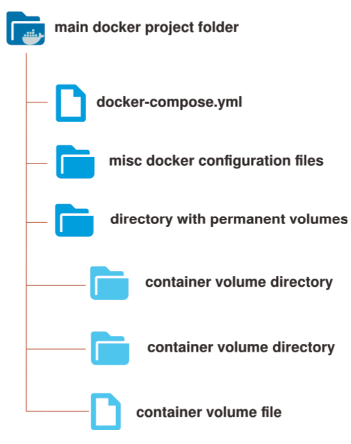


Figure 4 Organization of directories and files for managing a dockerized service in SEISMOFAULTS.EU.

Compose is a tool that Docker provides to manage applications that require multiple containers to run more easily. In our case, an OGC-type web service needs a server to publish the services (GeoServer) and a DBMS that manages the published databases to which GeoServer interfaces (PostgreSQL). The compose file describes all the parameters needed to start the containers, run them, and make them communicate with each other.

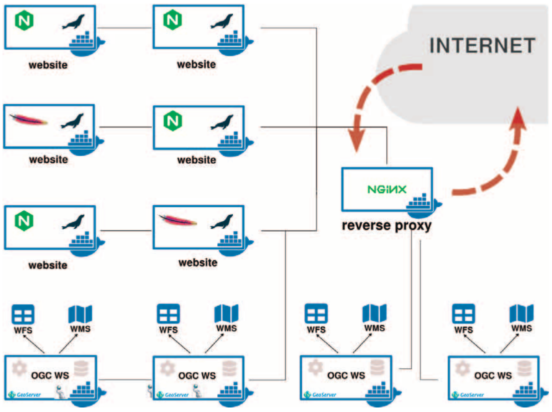


Figure 5 Docker containers scheme on SEISMOFAULTS.EU. Several websites and web services run on the same VM. Every website or web service includes its web server (Nginx or Apache), web application server (i.e., Apache Tomcat), and related RDBMSs. A (also "dockerized") reverse proxy manages network traffic between the infrastructure and the internet, routing arriving and departing connections to and from the suitable container hosting the service requested by the user.

We recommend reading the Docker (Table 1) official documentation for a complete understanding of how this technology works.
Service-related directories shown in Figure 4, are managed through Git. Git is a distributed version control system designed to handle software development projects. It allows multiple developers to work on the same project without overwriting each other's work. Each user has a local copy of the repository, making it easier to collaborate and work offline. Git tracks and manages changes to code, maintaining a complete history of modifications.
Every Docker project directory in SEISMOFAULTS.EU is also a Git repository. In this way, the Git system tracks every configuration change and synchronizes the sysadmin workstation and the server running the services straightforwardly, often automatically. Git repositories are hosted in the INGV GitLab service (https://gitlab.rm.ingv.it).

2. Security

All VMs are implemented with a strong rationalization of active services to limit unnecessary port exposure to the internet as far as possible.
Some VMs are available only from within the institutional intranet or are accessible only through the INGV Virtual Private Network. Those VMs publishing web services are protected by a few simple firewall rules.
Publicly exposed VMs are protected through Fail2ban (Table 1). Fail2ban is a software application designed to enhance system security by actively monitoring log files and automatically banning temporarily or blocking IP addresses that exhibit suspicious or malicious behavior (e.g., repeated connection attempts). It helps protect servers and services from various types of attacks, such as brute-force login attempts and distributed denial-of-service (DDoS) attacks, by dynamically adjusting firewall rules to deny access to potential threats.
Security administration policies also include, as commonly accepted good practices, software updates, periodical backup, and systems monitoring.

2.1 Software updates

The OS and application software updates are performed from some scripts that are automatically launched at predefined periods using standard Unix tools (basically, *Bash* shell scripts involving *cron* and *apt*). The system administrator performs a manual revision of updates on a weekly schedule.
The system administrator manually curates docker image updates and performs them when image distributors release security patches.

2.2 Backup

System backups are performed at several independent levels:

- VM snapshots are periodically made through the Hypervisor user interface, and the main ones are transferred in the form of Virtual Machine XVA files to external storage.
- Several scripts are executed during the least active hours to perform a backup of the most important directories of the VM OSs and of the container *volumes*, i.e., those container components where the containerized applications store data.

VM snapshots and backups are moved in disk partitions normally unused by the OS on different virtual devices on the SAN, linked through Network File System (NFS) protocol.

2.3 Monitoring and access statistics

Monitoring is the process of collecting, analyzing, and displaying usage data to measure the performance of an IT system. The goal is to ensure the system's availability, performance, and security. We monitor the "SEISMOFAULTS.EU" infrastructure to measure the performance of applications, databases, networks, and servers. Performance metrics such as latency, throughput, availability, and utilization provide a clear picture of the system's health and can identify potential areas of improvement to take corrective actions when needed. Monitoring can also provide insight into the security of the system. It can detect unauthorized access attempts, malware, and other malicious activity.
The SEISMOFAULTS.EU infrastructure monitoring is performed at multiple levels. At the first level, the VM hypervisor notifies the system administrator if a VM has an anomalous behavior, such as the excessive activity of the disks or CPUs for an extended time, a saturation of memory space, or abnormal network traffic (Figure 6).

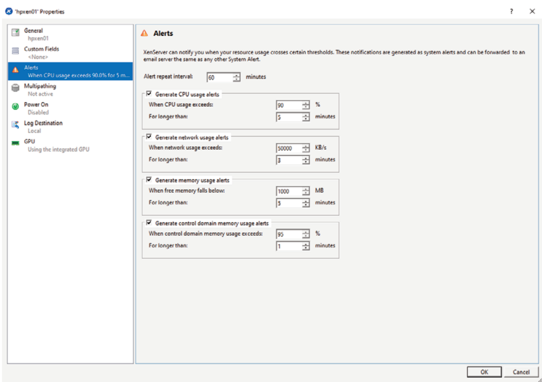


Figure 6 Hypervisor alert system for a server belonging to the SEISMOFAULTS.EU installation.

At the second level, Nagios Core (Table 1), a software specifically designed for this task configured on a dedicated VM, monitors the VMs at the OS level (processes, CPU load, RAM usage, disk partition usage, logged users, etc.) and at the process level (services and websites). Nagios is configured to send a notification to the system administrator through email or instant message in case of malfunction (Figures 7 and 8).

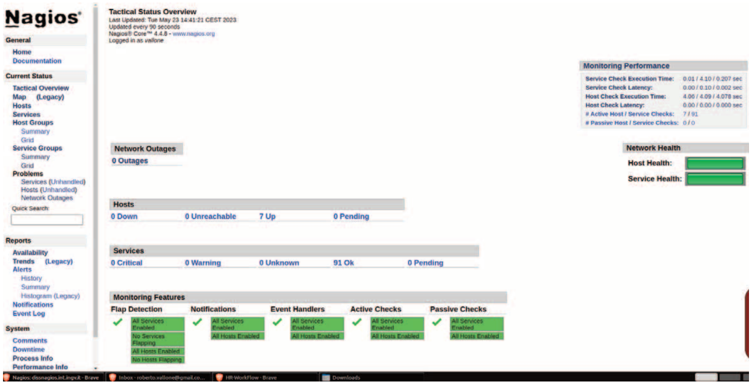


Figure 7 Nagios interface showing the so-called Tactical overview.

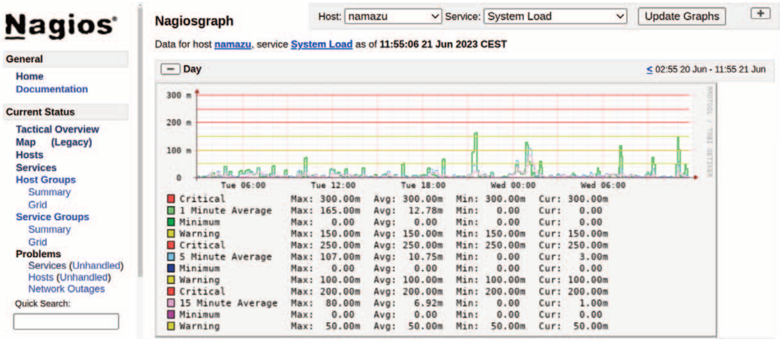


Figure 8 Nagios interface showing specific host performance.

SEISMOFAULTS.EU uses two approaches to monitor access and report usage statistics for websites and web services because of the different technologies underlying their operation. We create website access statistics through the Google Analytics platform and web services access statistics through the software AWStats (Table 1). Google Analytics works through a javascript that “intercepts” the connections to every webpage of the website and stores data in the Google Cloud for a wide range of analyses (Figure 9).

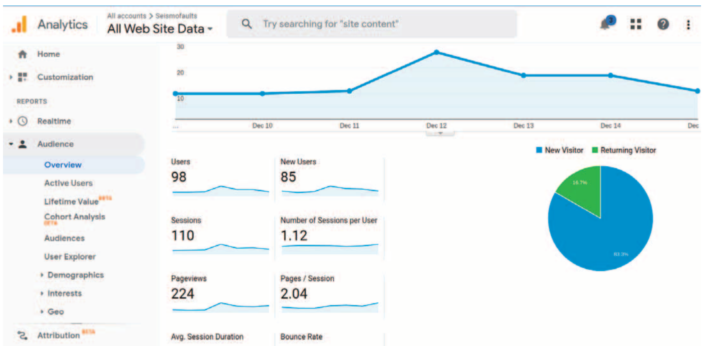


Figure 9 Google Analytics interface for tracking website traffic.

AWStats, instead, is a Perl script installed on the same VMs that publish the services. It analyzes the logs generated by the Web Application manager that runs GeoServer (see Table 1 and Chapter Functionality) and produces as output a webpage with a set of statistics (Figure 10).

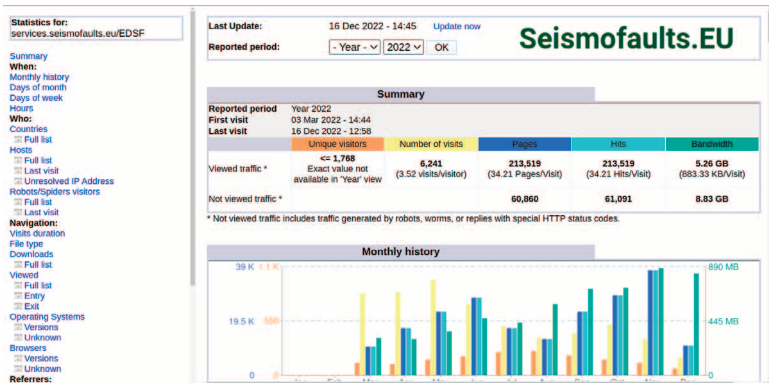


Figure 10 AWStats interface with access statistics to a web service (EDSF13 in this example).

3. Functionality

This Chapter describes the SEISMOFAULTS.EU organization for publishing datasets through web services.
All services are published from the primary node located in the Rome office, as well as all activities related to development, testing, monitoring and backup/recovery systems. The server hosted at the Bologna premises acts exclusively as a secondary node for publishing web services related to geographic datasets (which is then the main objective of the infrastructure). This node replaces the primary node in case of malfunctions that require recovery times longer than a few hours.
Appendix A provides a full list of websites and web services currently published by SEISMOFAULTS.EU.

3.1 Web Services publication

The combination of programs and applications to publish web services typically involves multiple software components that work together to serve geospatial data over the web. The components involved in the SEISMOFAULTS.EU infrastructure are summarized below.

1. Source Datasets: This is the source of the geospatial data, which can be originally implemented in various formats such as data tables, shapefiles, raster files, spatial databases, or other data repositories. Almost all data sources hosted by SEISMOFAULTS.EU are in the form of spatial databases. PostGIS is the open-source solution selected to manage such geographic data [Obe & Hsu, 2021].
2. Geospatial Data Processing and Management Software: It may include tools for data conversion, transformation, cleaning, and general data preparation tools. Examples of such software include Geospatial Data Abstraction Library [Rouault et al., 2023] and GIS software.
3. Geospatial Server Software: This software is the server component responsible for publishing the geospatial data as web services. This task is accomplished by the GeoServer software (Table 1).
4. Web Server Software: A web server is required to handle HTTP requests and responses.

It acts as an intermediary between the client applications and the geospatial server software. In our installation, this task is covered by a combination of two software: Apache Tomcat, which hosts the GeoServer application itself, and Nginx for connection proxying.

5. Web Service Protocols: The geospatial server software typically supports standard web service protocols such as the Web Map Service (WMS), Web Feature Service (WFS), or Web Coverage Service (WCS). These protocols define the rules and methods for accessing and interacting with the geospatial data.

3.1.1 PostGIS

PostGIS is an extension that adds spatial data capabilities to the PostgreSQL open-source relational database management system. Together, they provide a powerful solution for managing geographical databases. PostGIS extends PostgreSQL with spatial data types, spatial indexing, spatial functions, and operators for spatial queries, joins, and analyses. This combination of tools allows the users to efficiently store, retrieve, and analyze geospatial data within PostgreSQL, making it the most reasonable choice for the management of databases such as those dealt with by EDSF [Obe & Hsu, 2021].

3.1.2 Tomcat

Apache Tomcat is an open-source web server and servlet container known for its lightweight nature, high performance, and support for Java-based web applications. It is a platform for running servlets, JSPs, and Java EE technologies. Tomcat operates as a standalone web server, provides extensibility through additional components, offers security features, and includes management and monitoring capabilities. It is a popular choice for deploying dynamic and scalable Java web applications. In SEISMOFAULTS.EU Tomcat hosts the GeoServer application.

3.1.3 Nginx

Nginx is a high-performance, open-source web server and reverse proxy server known for its scalability, efficiency, and reliability. Originally designed to address the C10K problem [Kegel, 2011], Nginx has gained popularity as a versatile web server and a powerful reverse proxy solution. As a web server, Nginx efficiently serves static content, handles HTTP requests, and supports various web protocols. It can handle many simultaneous connections with minimal resource consumption, making it suitable for high-traffic websites.

In SEISMOFAULTS.EU Nginx is implemented as a reverse proxy. In this context, Nginx acts as an intermediary between clients (user browsers, web applications, GIS software) and backend servers (GeoServer hosting the requested databases). When a client sends a request, Nginx receives it and forwards it to the appropriate backend server based on defined rules. The backend servers are implemented inside Docker containers with Tomcat, GeoServer, and PostGIS.

Using Nginx as a reverse proxy offers several benefits:

1. Load Balancing: Nginx can distribute incoming requests across multiple backend servers, balancing the load and ensuring optimal resource utilization. It uses various load-balancing algorithms such as round-robin, least connections, IP hash, and more.
2. Caching: Nginx can cache frequently accessed static content, reducing the load on backend servers and improving overall performance. It supports both server-side and client-side caching mechanisms.

3. SSL Termination: Nginx offloads SSL/TLS encryption and decryption from backend servers, reducing their processing overhead. It can handle secure connections and perform SSL termination, providing secure communication between clients and backend servers.
4. High Availability and Failover: Nginx can ensure high availability and automatic failover by configuring multiple backend servers and implementing health checks. If a backend server becomes unavailable, Nginx can redirect requests to healthy servers, maintaining continuous service.
5. URL Manipulation and Rewriting: Nginx allows for URL manipulation and rewriting, enabling the modification of request URLs, redirections, and other routing operations. This flexibility provides options for URL-based routing, rewriting, and proxying.
6. Security Features: Nginx offers various security features, including access control, rate limiting, request filtering, and the ability to act as a reverse proxy firewall. It helps protect backend servers by filtering and blocking malicious requests.

Overall, Nginx's use as a reverse proxy enhances performance, scalability, and reliability in web application deployments. It efficiently distributes traffic, handles SSL/TLS encryption, caches content, and provides flexible routing and URL manipulation capabilities.

3.1.4 GeoServer

GeoServer is an open-source map server for managing, publishing, and sharing geospatial data using OGC protocols such as WMS and WFS (see the following section). GeoServer is written in Java and offers a wide range of functionalities that are summarized below.
GeoServer's functionalities for WMS include:

1. Map publication: GeoServer enables the publication of interactive maps based on geospatial data. It supports various data formats such as shapefiles, raster files, spatial databases, and many others.
2. Cartographic representation: GeoServer offers a wide range of options for customizing the appearance of published maps. Users can define symbols, colors, labels, and visualization styles for geospatial data.
3. Multiple layers: GeoServer allows the combination of different data layers to create composite maps. Layers of different formats and origins can be overlaid to create complex and informative maps.
4. Caching: GeoServer supports caching to improve performance and reduce server load. Map images can be stored in the cache for quick and redundant access.
5. Projection support: GeoServer handles a wide range of map projections and allows for the transformation of geospatial data between different projections. This enables the display of data in different coordinate reference systems and the combination of datasets from different sources.

GeoServer's functionalities for WFS include:

1. Vector data publication: GeoServer allows for the publication of vector geospatial data such as points, lines, and polygons via the WFS protocol. This data can be interactively queried and retrieved by clients.
2. Spatial querying: GeoServer supports advanced spatial queries to retrieve vector data based on their geographic location. Queries can be based on areas of interest, buffers, intersections, and other spatial operations.
3. Transactions: GeoServer supports the WFS-T (Web Feature Service - Transaction) protocol, which allows users to perform *insert*, *update*, and *delete* operations on vector

data within the server. Since datasets hosted by SEISMOFAULTS.EU pertain to published scientific sources, this service is not offered to the public.
4. Filters: GeoServer allows for the application of filters on the vector data published via WFS to limit results based on specific, user-selected criteria.

Overall, GeoServer provides a robust platform for publishing geospatial data and creating interactive maps through standard WMS and WFS protocols. With its open-source nature, GeoServer is widely used and supported by the geospatial community and was implemented in SEISMOFAULTS.EU since its beginning.

3.2 Open Geospatial Consortium web services

All EDSF datasets include intrinsic information on their geographical position, and are published through the standards developed by the Open Geospatial Consortium (OGC; https://ogc.org). OGC is an international consortium that develops open standards and specifications for geospatial data and technologies. The OGC provides a comprehensive web service suite that enables sharing and integrating geospatial data across different platforms and applications. The main standards developed by OGC and implemented in SEISMOFAULTS.EU are explained below.

3.2.1 Web Map Service (WMS)

The Web Map Service is a standard interface for serving georeferenced map images over the Internet. The WMS allows users to request maps of specific areas and layers that are dynamically generated by the server. The WMS supports a range of image formats, such as PNG, JPEG, and GIF. WMS is widely used in web mapping applications and is supported by most GIS software.

3.2.2 Web Feature Service (WFS)

The Web Feature Service is a standard interface for accessing and manipulating geospatial vector data over the internet. The WFS allows users to retrieve, insert, update, and delete vector features from a server. The WFS supports a range of data formats, such as GML and GeoJSON. WFS is widely used in GIS applications that require access to raw geospatial data.

3.2.3 Web Coverage Service (WCS)

The Web Coverage Service is a standard interface for accessing and retrieving geospatial coverages over the Internet. A coverage is a collection of geospatial data that represents a continuous field, such as a temperature or rainfall map. The WCS allows users to retrieve coverages as images or data files in a range of formats, such as NetCDF and GeoTIFF. WCS is widely used in scientific and environmental applications that require access to large amounts of geospatial data.

3.3 Benefits of OGC Web Services for EDSF data publication

Interoperability
The OGC web services provide a common set of interfaces and protocols that enable interoperability between different geospatial systems and applications. This means that data and services can be shared and integrated across different platforms, reducing duplication and increasing efficiency.

Standardization
The OGC web services are based on open standards and specifications that are developed and maintained by the consortium. This ensures that the services are widely adopted and supported by the geospatial community and that they are compatible with other standards-based systems and applications.

Flexibility
The OGC web services are designed to be flexible and extensible, allowing users to customize and adapt them to their specific needs. This enables users to create tailored solutions that meet their unique requirements without being limited by proprietary or closed systems.

The OGC web services are widely adopted and supported, making them a reliable and effective tool for creating innovative solutions and applications.

4. SEISMOFAULTS.EU and EPOS

The SEISMOFAULTS.EU infrastructure was created in the framework of the European Plate Observing System (EPOS) project since its Implementation Phase (EPOS-IP) in 2016 and, as part of the Thematic Corse Service (TCS) "Seismology" [Haslinger et al., 2022], participated in the formation of the EPOS European Research Infrastructure Consortium (EPOS-ERIC) in 2018.
SEISMOFAULTS.EU staff took part in the work for the implementation of the EPOS Data Portal officially presented to the scientific community during the European Geosciences Union conference held in Vienna in 2023 [Bailo et al., 2022]. Two datasets, at European scale, are presently incorporated in the EPOS Data Portal, the European Database of Seismogenic Faults [Basili et al., 2013] and the European Fault-Source Model [Basili et al., 2022] (Figure 11).

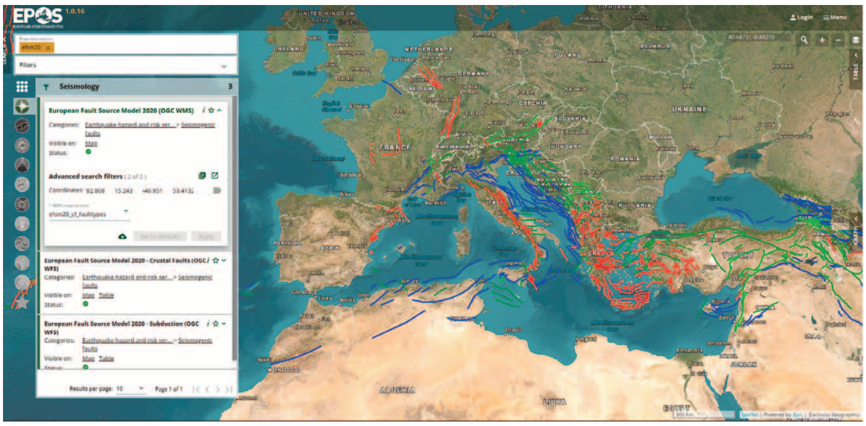


Figure 11 The European Fault-Source Model 2020 [Basili et al., 2022], published by SEISMOFAULTS.EU, as seen in the EPOS Data Portal.

Web services related to other datasets, such as the latest versions of the Database of individual Seismogenic Sources [DISS Working Group, 2015; 2018; 2021], are shared and developed in the framework of the Joint Research Unit of EPOS-Italia (https://www.epos-italia.it/).
SEISMOFAULTS.EU and EPOS maintain a collaborative relationship aimed at advancing seismological research and earthquake hazard analyses. Although they have distinct objectives, their shared goals and complementary capabilities make them mutually beneficial.

Data Sharing and Integration
SEISMOFAULTS.EU contributes its comprehensive database of seismogenic faults in Europe to the EPOS infrastructure. This integration allows users to seamlessly access fault-related data within the broader context of solid Earth science research. It enables researchers to combine fault information with other geophysical, geological, and geochemical data available through EPOS, fostering interdisciplinary studies.

Interoperability and Standardization
Both initiatives prioritize data harmonization, standardization, and the establishment of common metadata standards. This alignment facilitates the integration of fault data from SEISMOFAULTS.EU into the standard framework of EPOS. The interoperability ensures efficient data sharing, integration, and cross-disciplinary collaboration.

Enhanced Seismic Hazard Assessment
The partnership between SEISMOFAULTS.EU and EPOS.eu takes place within the TCS Seismology through the EFEHR consortium, thereby strengthening seismic hazard assessment capabilities. SEISMOFAULTS.EU provides detailed fault characterization, including fault geometry, slip rates, and paleoseismic data, which can be integrated with seismicity and other relevant data available through EPOS. This integration enhances the accuracy and completeness of seismic hazard models and contributes to more informed decision-making regarding earthquake risk management.

Research and Innovation
The partnership between SEISMOFAULTS.EU and EPOS.eu fosters research and innovation in seismology and related fields. By combining their resources, expertise, and networks, they create opportunities for collaborative projects (e.g., Geo-INQUIRE, https://www.geo-inquire.eu/; DT-GEO, https://dtgeo.eu/), knowledge exchange, and the development of advanced seismological models and methodologies. The joint efforts drive scientific advancements and contribute to the broader understanding of Earth's processes and seismic phenomena and meeting the new challenges of the exascale era of High-Performance Computing (HPC) [Folch et al., 2023].

Access to Infrastructure and expertise
SEISMOFAULTS.EU benefits from the distributed infrastructure and expertise provided by EPOS.eu. EPOS's established network of national research infrastructures and data centers offers technical support, data management capabilities, and access to computational resources. This infrastructure enables SEISMOFAULTS.EU to enhance its database management, data dissemination, and user accessibility, furthering its impact within the seismological community.

In conclusion, the relationship between SEISMOFAULTS.EU and EPOS.eu is characterized by collaboration, data integration, and a shared commitment to advancing seismological research. By leveraging their respective strengths and resources, these initiatives enhance seismic hazard assessment, promote interdisciplinary studies, and drive innovation in the field of seismology within the broader framework of Earth science research facilitated by EPOS.

5. GDPR compliance

SEISMOFAULTS.EU assures compliance with General Data Protection Regulation (https://gdpr-info.eu/) on all levels of the EDSF installation.
Since hosted data are distributed under Creative Commons licenses (https://creativecommons.org), there is no need to implement authentication systems, so there is no need to process or store personal data.
The hosted websites install no cookies on user browsers, except for the technical ones.
At the logging level, client IPs are anonymized immediately after the connection by a specific web server configuration rule and not stored elsewhere. Anonymized IPs are used for statistical and reporting purposes only.

6. Future developments

Even if the infrastructure has reached a satisfactory stage of maturity, there are still several improvements on the horizon.
A new client/server backup/restore platform is under implementation and testing, and it should be ready to be used by fall 2023.
By the end of 2023, the virtual machine architecture should be migrated from the current implementation with Xen Server to another (yet to be decided) virtualization platform.
Furthermore, to provide an increasingly reliability-oriented service, we are planning to configure a load-balancing service that is deployed in a cloud architecture and/or an orchestration platform like Kubernetes so as to implement high availability through the servers installed at the Rome and Bologna premises.

Acknowledgments

We thank all the INGV staff of Centro Servizi Informativi and particularly Giovanni Scarpato, Stefano Cacciaguerra, Pietro Ficeli, Manuela Sbarra, Gianpaolo Sensale, Diego Sorrentino, Stefano Vazzoler, Francesco Zanolin for their continuous IT support; the EPOS Integrated Core Services staff of the EPOS office in Rome: Daniele Bailo, Kety Giuliacci, Rossana Paciello, and Valerio Vinciarelli for their effort in EPOS/SEISMOFAULTS.EU cooperation; and our colleagues Valentino Lauciani for suggestions on the Docker configuration, Matteo Quintiliani for managing the INGV GitLab installation, and Giorgio Maria De Santis, Mario Locati, and Gabriele Tarabusi for the fruitful exchange of views.
The development of the SEISMOFAULTS.EU infrastructure benefitted from the funding of H2020 projects EPOS-IP (grant agreement No. 676564) and SERA (grant agreements No. 730900), the JRU EPOS-ITALIA - Piano di Attività 2021-2024 supported by the Italian Ministry of University and Research (MUR), and the DPC-INGV Agreement 2012-2021 (Annex A) and 2022-2024 (Annex A).
We also thank Stefano Cacciaguerra and Fabrizio Pistagna for their careful reviews which improved the quality of this report.

References

AGID, (2019). *Linee guida su acquisizione e riuso di software per le pubbliche amministrazioni*. Retrieved from https://www.agid.gov.it/sites/default/files/repository_files/lg-acquisizione-e-riuso-software-per-pa-docs_pubblicata.pdf

Anonymous, (2018). *DATA POLICY | EPOS*. Retrieved June 23, 2023, from https://www.epos-eu.org/node/91

Babeyko A., Lorito S., Hernandez F., Lauterjung J., Løvholt F., Rudloff A. et al., (2022). *Towards the new Thematic Core Service Tsunami within the EPOS Research Infrastructure*. Annals of Geophysics, 65(2), DM215. https://doi.org/10.4401/ag-8762

Bailo D., Michalek J., Jeffery K.G., Atakan K., Paciello R., & EPOS IT Team, (2022). *EPOS Data portal for cross-disciplinary data access in the Solid Earth Domain*. Presented at the EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-4265. https://doi.org/10.5194/egusphere-egu22-4265

Basili R., Brizuela B., Herrero A., Iqbal S., Lorito S., Maesano F.E. et al., (2021). *The Making of the NEAM Tsunami Hazard Model 2018 (NEAMTHM18)*. Frontiers in Earth Science, 8. https://doi.org/10.3389/feart.2020.616594

Basili R., Kastelic V., Demircioglu M.B., Garcia Moreno D., Nemser E.S., Petricca P. et al., (2013). *The European Database of Seismogenic Faults (EDSF) compiled in the framework of the Project SHARE*. Istituto Nazionale di Geofisica e Vulcanologia (INGV), https://doi.org/10.6092/INGV.IT-SHARE-EDSF. Retrieved from https://edsf13.ingv.it/

Basili R., Brizuela B., Herrero A., Iqbal S., Lorito S., Maesano F.E. et al., (2018). *NEAM Tsunami Hazard Model 2018 (NEAMTHM18): online data of the Probabilistic Tsunami Hazard Model for the NEAM Region from the TSUMAPS-NEAM project*. Istituto Nazionale di Geofisica e Vulcanologia (INGV); http://doi.org/10.13127/tsunami/neamthm18

Basili R., Danciu L., Beauval C., Sesetyan K., Vilanova S.P., Adamia S. et al., (2022). *European Fault-Source Model 2020 (EFSM20): online data on fault geometry and activity parameters*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/10.13127/EFSM20

DISS Working Group, (2015). *Database of Individual Seismogenic Sources (DISS), version 3.2.0*. A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas.

DISS Working Group, (2018). *Database of Individual Seismogenic Sources (DISS), version 3.2.1*. 127 Individual Seismogenic Sources, 188 Composite Seismogenic Sources, 35 Debated Seismogenic Sources, 3 Subduction Zones. https://doi.org/10.6092/INGV.IT-DISS3.2.1

DISS Working Group, (2021). *Database of Individual Seismogenic Sources (DISS), version 3.3.0*. A compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. 132 Individual Seismogenic Sources, 197 Composite Seismogenic Sources, 38 Debated Seismogenic Sources, 4 Subduction Zones. https://doi.org/10.13127/DISS3.3.0

Folch A., Abril C., Afanasiev M., Amati G., Bader M., Badia R.M. et al., (2023). *The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase*. Future Generation Computer Systems, S0167739X23001401. https://doi.org/10.1016/j.future.2023.04.006

Guidoboni E., Ferrari G., Mariotti D., Comastri A., Tarabusi G., Sgattoni G., & Valensise G., (2018). *CFTI5Med, Catalogo dei Forti Terremoti in Italia (461 a.C.-1997) e nell'area Mediterranea (760 a.C.-1500)* (Version 5), 1734 earthquakes, 43298 macroseismic Data Points, 2405 effects of the natural environment, 39924 historical critical comments, 47211 historical sources. https://doi.org/10.6092/INGV.IT-CFTI5

Guidoboni E., Ferrari G., Tarabusi G., Sgattoni G., Comastri A., Mariotti D. et al., (2019). *CFTI5Med, the new release of the catalogue of strong earthquakes in Italy and in the Mediterranean area*. Scientific Data, 6(1), 80. https://doi.org/10.1038/s41597-019-0091-9

Haslinger F., Basili R., Bossu R., Cauzzi C., Cotton F., Crowley H. et al., (2022). *Coordinated and Interoperable Seismological Data and Product Services in Europe: the EPOS Thematic Core Service for Seismology*. Annals of Geophysics, 65(2), DM213. https://doi.org/10.4401/ag-8767

Kegel D., (2011). *The C10K problem*. Retrieved June 16, 2023, from http://www.kegel.com/c10k.html

Obe R.O., & Hsu L.S., (2021). *PostGIS in action* (Third edition). Shelter Island: Manning Publications.

Puglisi G., Basili R., Chiodetti A.G., Cianchi A., Drudi M., Freda C. et al., (2016). *Principles of the INGV Data Policy | Principles of the INGV Data Policy*. Retrieved from https://data.ingv.it/docs/principles/

Rouault E., Warmerdam F., Schwehr K., Kiselev A., Butler H., Łoskot M. et al., (2023). *GDAL* (Version v3.7.0). Zenodo. https://doi.org/10.5281/ZENODO.5884351

APPENDIX A

Hosted websites

(to the date of this report)

Seismofaults.eu
    https://seismofaults.eu
Database of Individual Seismogenic Sources
    https://diss.ingv.it
European Database of Seismogenic Sources 2013
    https://edsf13.ingv.it
Tsunamidata.org
    https://tsunamidata.org
Hazard Curves (for NEAMTHM18 web services)
    https://hazardcurves.tsunamidata.org/
Tsumaps NEAM
    https://tsumaps-neam.eu
Interactive Tsunami Hazard Tool
    https://hazard.tsunamidata.org/
Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe (SERA) - Virtual Access 3 Task
    https://sera-va3.rm.ingv.it
INGV SismoLab-3D laboratory
    https://sismolab3d.ingv.it

APPENDIX B

Hosted web services

(to the date of this report)

European Database of Seismogenic Faults 2013
- OGC WMS
  - https://services.seismofaults.eu/EDSF/ows?service=WMS&request=GetCapablities
- OGC WFS
  - https://services.seismofaults.eu/EDSF/ows?service=WFS&request=GetCapablities

Database of Individual Seismogenic Sources v. 3.2.0
- OGC WMS
  - https://services.seismofaults.eu/DISS320/ows?service=WMS&request=GetCapabilities
- OGC WFS
  - https://services.seismofaults.eu/DISS320/ows?service=WFS&request=GetCapabilities

Database of Individual Seismogenic Sources v. 3.2.1
- OGC WMS
  - https://services.seismofaults.eu/DISS321/ows?service=WMS&request=GetCapabilities
- OGC WFS
  - https://services.seismofaults.eu/DISS321/ows?service=WFS&request=GetCapabilities

Database of Individual Seismogenic Sources v. 3.3.0
- OGC WMS
  - https://services.seismofaults.eu/DISS330/ows?service=WMS&request=GetCapabilities
- OGC WFS
  - https://services.seismofaults.eu/DISS330/ows?service=WFS&request=GetCapabilities

European Fault Source Model 2020
- OGC WMS
  - https://services.seismofaults.eu/EFSM20/ows?service=WMS&request=GetCapabilities
- OGC WFS
  - https://services.seismofaults.eu/EFSM20/ows?service=WFS&request=GetCapabilities

NEAM Tsunami Hazard Model 2018
- OGC WMS
  - https://services.tsunamidata.org/NEAMTHM18/ows?service=WMS&request=GetCapabilities
- OGC WFS
  - https://services.tsunamidata.org/NEAMTHM18/ows?service=WFS&request=GetCapabilities

Catalogo dei Forti Terremoti Italiani CFTI5Med
- OGC WMS
  - https://services-storing.ingv.it/CFTI/ows?service=WMS&request=getCapabilities
- OGC WFS
  - https://services-storing.ingv.it/CFTI/ows?service=WFS&request=getCapabilities

QUADERNI di GEOFISICA

ISSN 1590-2595 https://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941 https://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651 https://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

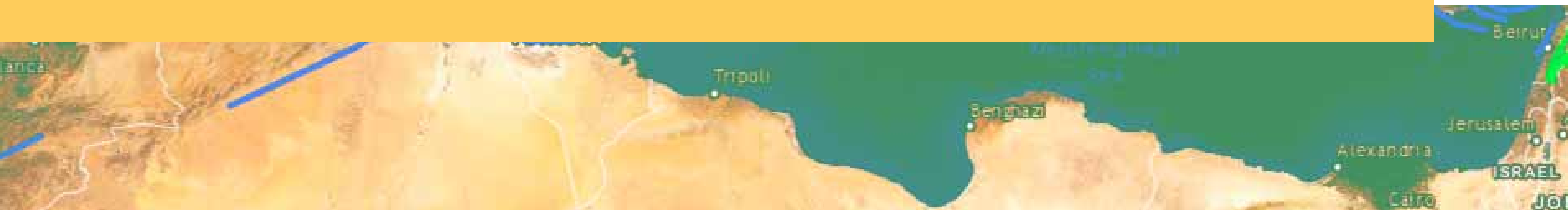
Coordinamento editoriale
Francesca DI STEFANO
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico
Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

Impaginazione
Barbara ANGIONI
Patrizia PANTANI
Massimiliano CASCONE
Istituto Nazionale di Geofisica e Vulcanologia

Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA