
Plan Overview

A Data Management Plan created using DMPonline

Title: European Fault-Source Model 2020 (EFSM20): online data on fault geometry and activity parameters

Creator: Roberto Basili

Principal Investigator: Roberto Vallone, Roberto Basili

Data Manager: Roberto Vallone, Roberto Basili

Affiliation: Other

Funder: European Commission

Template: DCC Template

ORCID iD: 0000-0003-1208-9412

ORCID iD: 0000-0002-1213-0828

Project abstract:

The [European Fault-Source Model 2020 \(EFSM20\)](#) is a product of the EU H2020 Project SERA (WP25-JRA3). It is designed to fulfill the requirements related to active faulting of the 2020 update of the European Seismic Hazard Model (ESHM20) following the probabilistic framework established for the 2013 European Seismic Hazard Model (ESHM13).

EFSM20 has two main categories of seismogenic sources: crustal faults and subduction systems. Crustal faults are meant to provide the hazard model with seismicity rates in various tectonic contexts, including onshore and offshore active plate margins and plate interiors. Subduction systems are intended to provide the hazard model with both slab interface and intraslab seismicity rates. The model covers an area encompassing a buffer of 300 km around all target European countries (except for Overseas Countries and Territories, OTCs) and a maximum of 300 km depth for slabs.

ID: 110029

Start date: 01-05-2017

End date: 30-04-2020

Last modified: 09-01-2025

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not

imply that the creator(s) endorse, or have any relationship to, your project or proposal

European Fault-Source Model 2020 (EFSM20): online data on fault geometry and activity parameters

Data Collection

What data will you collect or create?

The compilation of EFSM20 will exploit the wealth of information available from Earth Science studies, including, but not limited to, earthquake geology, seismology, seismotectonics, and geodynamics. The main purpose of such datasets is the geometric reconstruction of potential earthquake sources and estimating their activity rates. Re-used data will mainly come from the scientific literature.

The generated dataset will be an integrated data product from complex analyses or community-shared data harmonization.

The dataset will be compiled and distributed using open-source GIS software and open file formats. The data volume will be limited to a few megabytes which will pose no problem for long-term preservation and access.

How will the data be collected or created?

Generated and re-used data will most often be geospatial data providing the location of potential seismogenic faults, their geometry, and their behavior. Parameters detailing geometry and behavior will be linked to the spatial data in tabulated attributes. There is no community standard for this type of data; however, the dataset structure will follow prescriptions dictated by the needs of the earthquake hazard modelers.

For data available for download, there will be a different folder for each format. Each folder will be named with the dataset acronym followed by the format's name (e.g., EFSM20_GeoJSON). Each file will be named with the dataset acronym (EFSM20) followed by an abbreviation identifying the subset (e.g., "CF" for crustal faults) and the proper format extension (e.g., ".geojson"). Details about the naming rules will be given in the dataset documentation. The same naming rules will be applied to the OGC web service layers.

Once the dataset is published, there will only be one version of its final release. Additional versions are not planned. If an update becomes necessary, a different DOI will identify it.

The quality control of the distributed data will be carried out according to a multi-step workflow described in the [data quality assurance](#) document available in the documentation section of the [EDSF portal](#).

Documentation and Metadata

What documentation and metadata will accompany the data?

The EFSM20 dataset will be accompanied by comprehensive documentation addressing the data structure, the definition of variables, and the units of measurement.

Metadata will be openly available and contain enough information (direct link) to enable the user to access the data.

Provisions for metadata will include:

- metadata offered with the DOI as required by [DataCite](#);
- metadata offered through the [INGV Open Data Portal](#);
- metadata offered through the standard OGC protocol [CSW](#);
- EPOS-DCAT-AP when the dataset will be mapped in the [EPOS ICS-C portal](#);
- INSPIRE if the dataset will be mapped in the Italian ["Repertorio Nazionale dei Dati Territoriali"](#).

Ethics and Legal Compliance

How will you manage any ethical issues?

There is no ethical reason that could impact data distribution and sharing. A disclaimer will be associated with the dataset to remove legal liability from the data owner and publisher. Users will also be cautioned to consider the nature of the dataset carefully before using it for decisions concerning personal or public safety or business involving substantial financial or operational consequences.

No personal data will be collected or distributed with the dataset.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

The EFSM20 dataset will be distributed under the [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) license terms. Users can request additional permissions to use the dataset by [contacting the persons indicated on the website](#)

Storage and Backup

How will the data be stored and backed up during the research?

The data will be stored in the server that publishes the static file and in the server that issues the OGC services.
The data will be backed up using a storage server connected to the INGV private network.
To back up the database, we will use the standard PostgreSQL tool "pg_dump."
Since EFSM20 will be a single-version release and will not be updated, there is no need to schedule an automatic backup procedure.
The entire website where EFSM20 is published is regularly backed-up.
The responsible for the backup and recovery procedure is Roberto Vallone (INGV).
In case of an incident with the publishing server, data will be recovered by restoring the database and the files from one of the multiple backup services. In particular, the database will be restored using the standard "pg_restore" tool of PostgreSQL.

How will you manage access and security?

All EFSM20 data will be openly accessible.
SSL transfer for HTTP (HTTPS) is implemented and is chosen per default for all hosted services on the [EDSF Installation](#) where EFSM20 is published.
No sensitive data will be stored.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

EFSM20 data and metadata stored in the INGV repositories will remain available indefinitely.
EFSM20 is an integrated data product; as such, all the raw and processed data used to compile the EFSM20 dataset will remain with their owners.
EFSM20 will initially be used to devise the input dataset for the 2020 update of the European Seismic Hazard Model. In the future, EFSM20 will likely be used to carry out earthquake hazard analyses (e.g., ground shaking or tsunami), earthquake scenarios, or seismotectonic and geodynamic models.

What is the long-term preservation plan for the dataset?

The datasets will be deposited in two INGV servers installed on two different institutional premises for security reasons. Since the EFSM20 main dataset and envisaged derived products should not occupy more than 1 GB of disk storage and the file formats will presumably remain common for many years, the storage cost can be considered negligible.

Data Sharing

How will you share the data?

The standard OGC protocols WMS and WFS will be adopted to guarantee interoperability with other datasets or spatial data. The EFSM20 datasets will be available as downloadable files in popular formats (GeoJSON files, ESRI shapefiles, MapInfo Tables), facilitating users' combining and analyzing EFSM20 with other geographically referenced data in a desktop Geographic Information System (GIS).

The already reserved DOI <https://doi.org/10.13127/efsm20> will permanently identify the EFSM20 main dataset. Child DOIs (e.g.,...efsm20/"*derived dataset name*") will identify additional derived datasets or products.

Are any restrictions on data sharing required?

The entire EFSM20 dataset will be made openly accessible with no restrictions except for properly using the citation prescribed by the attribution license.

Responsibilities and Resources

Who will be responsible for data management?

The persons responsible for the data management, curation, preservation, and distribution are the [contact persons](#) that will be indicated on the website.

What resources will you require to deliver your plan?

Storage, archiving, re-use, and security costs will partly be covered by EPOS and INGV institutional funding. When additional resources are necessary, they will be sought through project funding.

5