Plan Overview

A Data Management Plan created using DMPonline

Title: European Database of Seismogenic Faults 2013

Creator: Roberto Vallone

Principal Investigator: roberto.basili@ingv.it, Roberto Basili

Data Manager: roberto.basili@ingv.it

Affiliation: Other

Funder: European Commission

Template: DCC Template

ORCID iD: 0000-0002-1213-0828

Project abstract:

The European Database of Seismogenic Faults 2013 (EDSF13) was compiled in the framework of the EU Project SHARE, Work Package 3, Task 3.2. EDSF13 includes only faults deemed capable of generating earthquakes of magnitude equal to or larger than 5.5 and aims at ensuring a homogeneous input for use in ground-shaking hazard assessment in the Euro-Mediterranean area. Several research institutions participated in this effort with the contribution of many scientists (see the EDSF13 web page for a full list). The EDSF13 dataset and website are hosted and maintained by INGV.

ID: 23124

Last modified: 16-11-2022

European Database of Seismogenic Faults 2013

Data Collection

What data will you collect or create?

The following data will be collected: data about the geology and active tectonics of an area stretching along and away from the main plate tectonic boundary between Europe and Africa, from west of the Gibraltar Straits to the easternmost part of Anatolia and from the Atlas to the south and the Lower Rhine Embayment to the north. 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 an acronym identifying the specific dataset followed by the format's name (e.g., ESRI_shapefile). Each file will be named with the specific dataset name (e.g., Crustal_fault_sources) followed by the proper format extension (e.g., ".shp"). Details about the content of each specific dataset will be provided in a README file added to each file folder. Specific 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 thedata quality assurance document available in the documentation section of the EDSF portal.

Documentation and Metadata

What documentation and metadata will accompany the data?

The EDSF13 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 byDataCite;
- 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 theEPOS 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 carefully consider the dataset's nature 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 EDSF13 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 EDSF13 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 EDSF13 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 EDSF13 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 EDSF13 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?

EDSF13 data and metadata stored in the INGV repositories will remain available indefinitely.
EDSF13 is an integrated data product; as such, all the raw and processed data used to compile the EDSF13 dataset will remain with their owners.
EDSF13 will initially be used to devise the input dataset for the first release of the European Seismic Hazard Model. In the future, EDSF13 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?

Datasets will be deposited in two servers owned by INGV, installed in two different institutional premises for security reasons. Since the EDSF13 dataset will occupy less than 10 MB of disk storage and the file formats used will presumably be of common use for many years, the cost of storage related 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 availability of the EFSM20 datasets as downloadable files in popular formats (GeoJSON files, ESRI shapefiles, MapInfo Tables) will facilitate users combining and analyzing EFSM20 with other geographically referenced data in a desktop Geographic Information System (GIS).
The already reserved DOI http://doi.org/10.6092/INGV.IT-SHARE-EDSF will permanently identify the EDSF13 dataset.

Are any restrictions on data sharing required?

The entire EDSF13 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 thecontact 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.