Plan Overview

A Data Management Plan created using DMPonline

Title: Monitoring Earth’s Evolution and Tectonics, WP11 - IPSES - Italian Platform for Solid Earth Science, Activity 11.9a - Services and interoperability layers for distributing earthquake faulting data in 4D

Creator: Roberto Basili

Principal Investigator: Roberto Basili

Data Manager: Roberto Basili, Pierfrancesco Burrato, Francesca Romana Cinti, Roberto Vallone

Project Administrator: Roberto Basili

Affiliation: Other

Template: DCC Template

ORCID iD: 0000-0002-1213-0828

Project abstract:

MEET is a project coordinated by INGV and funded by PNNR. It will allow to renew, implement and in some cases create monitoring nodes to build networks of research infrastructures fit for the future, in line with European and international standards.

By improving data production and integration, MEET aims to contribute to increasing awareness of Earth dynamics that will be crucial in the coming decades.

National Recovery and Resilience Plan (PNRR) - Mission 4, “Education and Research” - Component 2, “From research to business” - Investment line 3.1, “Fund for the creation of an integrated system of research and innovation infrastructures” - project code IR0000025

Activity 11.9a description
Earthquake faulting is a complex process encompassing different spatio-temporal scales. Datasets of active faults and seismogenic sources typically provide one or just a few facets at a time of this many-sided system. This activity aims to coordinate the geological and geophysical communities, particularly those working on earthquake hazards (e.g., surface displacement, ground failures, ground-shaking, tsunami), to develop new datasets and strengthen existing ones to provide innovative, strategic data-service portfolios.
After establishing the main requirements, we will design workflows for the following data: 3D fault geometries of seismogenic sources with constraints from geological 3D models (see OU 15 ISPRA); connections between 3D fault geometries and 2D fault surface traces (see OU 15 ISPRA); integration between spatial parameters of coseismic surface ruptures and earthquake-induced ground failures (see OU 15 ISPRA); paleoearthquakes data. The 4th dimension will be ensured by integrating the timing of modern and ancient surface ruptures and ground failures.
The development of these new datasets will be aided by the Geology and Geotechnologies Laboratory (https://bit.ly/33nPfCj) and the Reflection Seismology Laboratory “SismoLab-3D” (https://sismolab3d.ingv.it/). Coordination meetings will be organised to harmonise the

different contributions.
Web services of 4D interoperable data and metadata, using EPOS-compliant protocols (considering persistent identifiers, licences, metadata, documentation, and DMP), will be distributed through the SEISMOFAULTS.EU Platform (https://www.seismofaults.eu/) and linked to the Italian Platform for Solid Earth Science. Users will thus be enabled to explore and retrieve information from traditionally disconnected datasets. For example, it will be possible to integrate active fault data from different sources, improving the findability and accessibility of such information by a wide community of users and stakeholders.

ID: 175506

Start date: 01-11-2022

End date: 31-10-2025

Last modified: 06-05-2025

Grant number / URL: IR0000025 https://meet.ingv.it/

Copyright information:

Monitoring Earth’s Evolution and Tectonics, WP11 - IPSES - Italian Platform for Solid Earth Science, Activity 11.9a - Services and interoperability layers for distributing earthquake faulting data in 4D

Data Collection

What data will you collect or create?

The Activity 11.9a will collect data on earthquake faulting.
Four main datasets will constitute a portfolio of interconnected products.

- SF3D - Seismogenic Faults in Three Dimensions;
- SFIT - Surface Faulting events of ITaly;
- DIP - Database of Italian Paleo Earthquake;
- Interoperability layer between seismogenic sources of the DISS database and active and capable faults of the ITHACA database

How will the data be collected or created?

SF3D - Seismogenic Faults in Three Dimensions.
The 3D fault reconstructions will exploit original and published data on fault geometry at seismogenic depth. The data will be extracted from geological maps, geological sections, paleoseismic trenches, seismic reflection profiles, exploration wells, seismic tomography, location of earthquake sequences, and seismological analyses. A dedicated workflow and software tool will be developed to store and process the fault location data and generate 3D meshes made of triangular elements.

SFIT - Surface Faulting events of ITaly.
The SFIT database will include the information on the Surface Faulting produced by ITalian earthquakes derived from field and remote reconnaissance following the earthquake. The data and images will be derived uniquely from published scientific papers, and they will be parsed to be inserted into the Database. The relevance of SFIT will be to have all the available information describing in detail the surface ruptures organised and harmonised in a single structure.

DIP - Database of Italian Paleo Earthquake.
The DIP database will contain paleoseismic data collected at specific sites along the active faults in Italy. The recognition and ages of paleoearthquakes from trenches, caves, quarries, and exhumed fault planes will be derived from published scientific papers and will be parsed in order to populate the database structure. The relevance of DIP will be to have all the available information harmonised in a single database, facilitating the processing and the temporal and spatial analysis of the fault ruptures.

Interoperability layer between seismogenic sources of the DISS database and active and capable faults of the ITHACA database
The interoperability layer will be created by exploring the possible relationships between the composite seismogenic sources of the DISS database (CSS) and the active and capable faults of the ITHACA database (FAC). The data from the two databases will be retrieved using their web services. A dedicated informatics workflow exploiting published empirical relationships will be developed to

highlight the probability of reactivation of each individual FAC following scenario earthquakes that ruptured segments of the CSSs. The empirical relationships that will be used give the probability of reactivation as distributed faulting or primary faulting as a function of the minimum distance of the FAC from the surface trace of the CSS and are distinguished on the basis of the fault kinematics. Each individual FAC will be given a total probability of reactivation as the sum of the probabilities of being reactivated by each surrounding CSS segment. To ensure that the interoperability layer always uses the latest available contents of the two databases, which are released with different frequencies, and that the relationships between CSSs and FACs will always be updated, the informatics workflow will run at specific time intervals (e.g. once a month).

Documentation and Metadata

What documentation and metadata will accompany the data?

The "Fault data in 4D" datasets 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 links) 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.

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 "Fault data in 4D" datasets will be distributed under theCreative Commons Attribution 4.0 International (CC BY 4.0) license terms. Users can request additional permissions to use the datasets 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."

Static datasets: since successive dataset releases are not yet scheduled, there is no need to schedule an automatic backup procedure. A backup procedure will be run when newer versions become available.

Dynamic datasets: A backup procedure will be scheduled according to the frequency of data updates.

The entire website where "Fault data in 4D" datasets are 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 datasets 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 the datasets are 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?

"Fault data in 4D" data and metadata stored in the INGV repositories will remain available indefinitely.

"Fault data in 4D" datasets are integrated data products; as such, all the raw and processed data used to compile the various datasets will remain with their owners.

"Fault data in 4D" datasets will initially be used to devise possible input datasets 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?

For security reasons, the datasets will be deposited in two INGV servers installed on two different institutional premises. Since the main datasets and envisaged derived products should not occupy

more than 100-200 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 protocol WFS and WMS will be adopted to guarantee interoperability with other datasets or spatial data.

The datasets will be available as downloadable files in the GeoJSON open format, facilitating users' combining and analyzing the datasets with other geographically referenced data in a desktop Geographic Information System (GIS).

A DataCite DOI (https://doi.org/10.13127/****) will permanently identify all the dataset's first versions. DOIs that include the version number will identify subsequent updates.

Are any restrictions on data sharing required?

All datasets 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?

EPOS and INGV institutional funding will partly cover storage, archiving, re-use, and security costs after the project's end. When additional resources are necessary, they will be sought through project funding.