Plan Overview

A Data Management Plan created using DMPonline

Title: Engineering Strong Motion Database

Creator: Lucia Luzi

Data Manager: Claudia Mascandola, Chiara Felicetta, Emiliano Russo

Project Administrator: Lucia Luzi

Contributor: Claudia Mascandola, Sara Sgobba, Francesca Pacor, Lucia Luzi, Giovanni Lanzano

Affiliation: Other

Funder: European Commission

Template: Horizon 2020 DMP

Project abstract:

The Engineering Strong Motion database (ESM) is the collection of the accelerometric recordings of earthquakes with magnitude larger than 4 in the Euro-Mediterranean region, since 1969. ESM is designed to provide end users with quality-checked, uniformly processed strong-motion data and relevant parameters. The database is intended for various stakeholders (expert seismologists, earthquake engineers, students, and professionals) and has a user-friendly and straightforward web interface. Specific tools are available to users to process strong-motion data and select ground-motion suites for code-based seismic structural analyses. Data and metadata can be accessed through web services

ID: 23829

Start date: 01-01-2010

Last modified: 28-11-2024

Engineering Strong Motion Database - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

The collection of accelerometric records in an organized archive is useful to provide end users with quality-checked, uniformly processed strong-motion data and relevant parameters for several applications, such as predicting ground shaking and performing structural seismic analysis.

The earthquake recordings are distributed as ASCII, miniseed and HDF5 files. Unprocessed acceleration, processed acceleration, velocity and displacement time series and acceleration and displacement response spectra (5% damping) can be downloaded with relevant strong motion metadata.

The origin of earthquake recordings is the European Integrated Data Archive (EIDA). Data not available in the EIDA platform are obtained from original open-access databases:

1) Unified HEllenic Accelerogram Database (HEAD), containing Greek waveforms and metadata from 1973 to 1999;

2) ITACA, the database of Italian strong-motion data;

3) TR-NSMN, containing the Turkish dataset from 1976 to 2007;

4) Pan-European data from 1972 to 2008, not included in regional databases, are extracted from the European Strong Motion Database.

The Engineering Strong-Motion database is daily updated and the expected rate of growth is 300Gb per year

2. FAIR data

2.1 Making data findable, including provisions for metadata:

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

The earthquake recordings in the Engineering Strong Motion database can be discovered by specifying 35 metadata related to stations, events, or waveforms.

The ESM database can be uniquely identified through a Digital Object Identifier (10.13127/ESM). DOIs are also available to identify the single contributing networks.

The SEED convention is used to name Network codes, Station codes, Location IDs and Channel codes.

The standard used for metadata is Data Cite

2.2 Making data openly accessible:

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

In the Engineering Strong Motion database a period of embargo can be applied to earthquake records. In particular, records are restricted until the processed version is available (usually it takes 1 day to a week) or in case of protection of data of temporary experiments (e.g. Alparray project, seismic microzonation of the area affected by the 2016 central Italy sequence).

Data are made available in different formats:

1) parametric table (flatfile) available at https://esm-db.eu/#/products/flat_file; it is associated with a DOI (**https://doi.org/10.13127/esm/flatfile.1.0)** and it is distributed with the licence CC-BY

2) waveforms are available as ASCII, miniseed and HDF5 files at https://esm-db.eu/. Different licenses can be associated with the records, depending on the data provider's policy. Waveforms are also available at the webservice https://esm-db.eu/esmws/eventdata/1/

3) event metadata are available at the webservice https://esm-db.eu/fdsnws/event/1/ and the web site HTTP://esm-db.eu

4) station metadata are available at the webservice https://esm-db.eu/fdsnws/station/1/ and the web site HTTP://esm-db.eu

5) USGS shakemap input is available in .XML format, through the webservice http://esm.mi.ingv.it/esmws/shakemap/1/ for single earthquakes.

The entire database is associated with the CC-BY licence

2.3 Making data interoperable:

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Data interoperability is guaranteed by adopting the seismological standards of the International Federation of Digital Seismograph Networks (FDSN).

The vocabulary for all data types present in the Engineering Strong-Motion database is built in the framework of EPOS.

2.4 Increase data re-use (through clarifying licenses):

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Licenses are associated with the products distributed by ESM with a different degree of granularity.

The entire database has the CC-BY license

Single records have licenses that depend on the data provider policy (e.g. CC-BY-NC or CC-BY)

The USGS shakemap .xml has no license

The parametric table is associated with CC-BY

Data are assumed to be available and re-usable for the duration of the EPOS infrastructure

3. Allocation of resources

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

The cost to make available the data of the Engineering Strong Motion database is about 200 keuro/year. These resources come from INGV in-kind contributions, Italian and European projects, the ORFEUS consortium, the Italian Civil Protection and the JRU-EPOS Italy.

The long-term preservation of the infrastructure depends on INGV in-kind contributions, the civil protection funds and the resources that will be allocated by ORFEUS and the Italian government through EPOS Italy.

The data management is in charge of the ESM working group and Strong-Motion Management Committee of ORFEUS.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Secure data storage of the Engineering Strong-Motion database is assured by daily backups performed on a server located at INGV Milan and on a mirror at a server located at the Rende (CS) INGV department.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

At the moment no consent for data sharing and long-term preservation is included in questionnaires dealing with personal data

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

none

Engineering Strong Motion Database - Detailed DMP

1. Data summary

State the purpose of the data collection/generation

The purpose of data collection is to provide end users only with quality-checked, uniformly processed strong-motion data and relevant parameters for several applications, such as predicting ground shaking and performing structural seismic analysis.

Explain the relation to the objectives of the project

The Engineering Strong-Motion database is one of the services provided by ORFEUS, which, in turn, is one of the pillars of the EPOS seismology, one of the Thematic Core Services of EPOS

Specify the types and formats of data generated/collected

Data are made available in different formats:
1) parametric table (flatfile) available at http://esm.mi.ingv.it/flatfile-2017/flatfile.php upon user registration. It is a comma delimited file containing the relevant strong-motion parameters, the earthquake and station metadata associated to each record contained in the database.
2) earthquake records as ASCII files (unprocessed and processed acceleration, velocity and displacement time-series and acceleration and displacement response spectra with 5% damping)
3) USGS shakemap xml files to be used as input for the USGS shakemap software

Specify if existing data is being re-used (if any)

The records contained in the database are re-used from existing open archives and databases:
1) European Integrated Data Archive (EIDA)
2) Incorporated Research Institutions for Seismology (IRIS)
3) Unified HEllenic Accelerogram Database (HEAD)
4) ITACA the database of Italian strong-motion data
5) TR-NSMN containing the Turkish dataset from 1976 to 2007
6) European Strong Motion Database, pan-European data from 1972 to 2008
7) Road, Housing and Urban Development Research Center (BHRC), Iran, after a memorandum of understanding between BHRC and ORFEUS (Observatories & Research Facilities for European Seismology)

Specify the origin of the data

Earthquake recordings are mainly obtained from the European Integrated Data Archive (EIDA). Data not available in the EIDA platform are taken from original databases. Three main regional databases are identified:

1) Unified HEllenic Accelerogram Database (HEAD, see Data and Resources), released in 2004 and containing Greek waveforms and metadata from 1973 to 1999 (Theodulidis et al., 2004)

2) ITACA (see Data and Resources), the database of Italian strong-motion data from 1972 to 2015 (Luzi et al., 2008; Pacor, Paolucci, Luzi, et al., 2011);

3) TR-NSMN (see Data and Resources), containing the Turkish dataset from 1976 to 2007 (Akkar et al., 2010);

Pan-European data from 1972 to 2008, not included in regional databases, are extracted from the European Strong Motion Database (http://www.isesd.hi.is/).

Iranian data are obtained from the Road, Housing and Urban Development Research Center (BHRC) after a memorandum of understanding between BHRC and ORFEUS (Observatories & Research Facilities for European Seismology)

State the expected size of the data (if known)

The Engineering Strong-Motion database is daily updated and the expected rate of growth is 200Gb per year

Outline the data utility: to whom will it be useful

The database is designed for a large variety of stakeholders, such as expert seismologists, earthquake engineers, students, and professionals.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

In the web site of the Engineering Strong-Motion database the earthquake recordingscan be discovered by specifying 35 parameters related to stations, events, or waveform metadata. These metadata are defined internally.

The standard used for metadata used to discover ESM is Data Cite

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

The ESM database can be uniquely identified through a Digital Object Identifier (10.13127/ESM).

The parametric table (flatfile), available at http://esm.mi.ingv.it/flatfile-2017/flatfile.php, is associated

to the persistent identifier (PID) 11099/ESM_6269e409-ea78-4a00-bbee-14d0e3c39e41_flatfile_2017
Data providers are alwas associated to a unique IDs. Data providers are responsible of their DOIs

Outline naming conventions used

The earthquake recordings in ESM are named using the SEED convention. In particular, the SEED standards are used to name Network codes, Station codes, Location IDs and Channel codes.

Outline the approach towards search keyword

The search keywords have been identified considering the most used seismic event, recording station or waveform metadata (e.g. magnitude, date and time, country code, geographic constraints, intensity measure, event and station unique identifiers, etc)

Outline the approach for clear versioning

Software clear versioning is ensured by the use of gitlab.rm.ingv.it, so multiple teams can work together to create the product.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

The standard used for metadata used to discover ESM is Data Cite

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

In the Engineering Strong-motion database earthquake records are open, as well as the relevant strong-motion parameters and station and events metadata.
A period of embrago can be applied to earthquake records. In particular, records ar restricted until the processed version is available (usually it takes from 1 day to a week), or when there is the need of protect data of temporary experiments (e.g. Alparray project, seismic microzonation of the area affected by the 2016 central Italy sequence).

Specify how the data will be made available

Data are made available in different formats:

1) parametric table (flatfile) available at http://esm.mi.ingv.it/flatfile-2017/flatfile.php upon user registration
2) ASCII files available at http:\\esm.mi.ingv.it upon user registration
3) USGS shakemap xml, through a the webservice http://esm.mi.ingv.it/esmws/shakemap/1/

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

No software is needed to access the data. Data can be directly accessed from the ESM web site or through webservices

Specify where the data and associated metadata, documentation and code are deposited

Data (waveforms and database) are stored at INGV Milano and a backup in the Rende Department of INGV. The codes are deposited in https://gitlab.rm.ingv.it/

Specify how access will be provided in case there are any restrictions

In case there are restrictions no exceptions are made, so no data access is permitted

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Data are interoperable with other disciplines through the EPOS portal https://www.epos-eu.org/dataportal.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

The data are interoperable through seismological standards, deposited at the International Federation of Digital Seismograph Networks and OGC standards.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

The entire ESM database is licensed as CC-BY-NC-ND, in order to avoid duplication of the database
The parametric table (flatfile), available at http://esm.mi.ingv.it/flatfile-2017/flatfile.php, is distributed with the licence CC-BY-NC
The earthquake records, distributed as ASCII files, available at http:\\esm.mi.ingv.it, are associated to different licenses, depending on the data provider policy
The USGS shakemap xml is not currently associated toa license

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Data are made available for re-use after the records are manually processed by an ESM operator. This may take from 1 day to 1 week.
A longer period of embargo can be applied for sensitive data (e.g. specific projects) and depends on the rules of the specific data collection

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Data can be re-used after they are made available, until the ESM database will be maintained by EPOS

Describe data quality assurance processes

Data quality is assured by applying standard processing techniques, recognised by the scientific community.

Specify the length of time for which the data will remain re-usable

The time period for data re-use depends on EPOS

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

The cost for making the data of the Engineering Strong-Motion database available is about 200

keuro/year. These resources comes from INGV inkind contribution and Italian and European projects (included EPOS-IP).
The long time preservation of the infrastructure depends on the inkind contribution and the resorces that will be allocated from the Italian government through EPOS.

Clearly identify responsibilities for data management in your project

ESM is maintained at INGV (responsible Lucia Luzi).
ESM is distributed under the umbrella of ORFEUS, which, in turn, is one of the pillars of EPOS seismology, one of the EPOS Thematic Core Services. The management of the acceleration and strong-motion data in ORFEUS is curated by a Service Management Committee, composed by the representatives of the networks contributing to the database. A panel of external experts periodically gives user feedbacks, in order to match the needs of data users.

Describe costs and potential value of long term preservation

Long term preservation of the ESM database has a cost of 200 keuro /year, including the personnel dedicated to the data processing and the metadata revision, the hardware and software maintainance, the data curation and preservation.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

A daily backup of the ESM database is carried out in one NAS server at INGV.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

No ethical or legal issues are involved.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

None

Engineering Strong Motion Database - Final review DMP

1. Data summary

State the purpose of the data collection/generation

Question not answered.

Explain the relation to the objectives of the project

Question not answered.

Specify the types and formats of data generated/collected

Question not answered.

Specify if existing data is being re-used (if any)

Question not answered.

Specify the origin of the data

Question not answered.

State the expected size of the data (if known)

Question not answered.

Outline the data utility: to whom will it be useful

Question not answered.

2.1 Making data findable, including provisions for metadata [FAIR data]

Outline the discoverability of data (metadata provision)

Question not answered.

Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?

Question not answered.

Outline naming conventions used

Question not answered.

Outline the approach towards search keyword

Question not answered.

Outline the approach for clear versioning

Question not answered.

Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how

Question not answered.

2.2 Making data openly accessible [FAIR data]

Specify which data will be made openly available? If some data is kept closed provide rationale for doing so

Question not answered.

Specify how the data will be made available

Question not answered.

Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?

Question not answered.

Specify where the data and associated metadata, documentation and code are deposited

Question not answered.

Specify how access will be provided in case there are any restrictions

Question not answered.

2.3 Making data interoperable [FAIR data]

Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.

Question not answered.

Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

Question not answered.

2.4 Increase data re-use (through clarifying licenses) [FAIR data]

Specify how the data will be licenced to permit the widest reuse possible

Question not answered.

Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed

Question not answered.

Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why

Question not answered.

Describe data quality assurance processes

Question not answered.

Specify the length of time for which the data will remain re-usable

Question not answered.

3. Allocation of resources

Estimate the costs for making your data FAIR. Describe how you intend to cover these costs

Question not answered.

Clearly identify responsibilities for data management in your project

Question not answered.

Describe costs and potential value of long term preservation

Question not answered.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

Question not answered.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

Question not answered.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Question not answered.