Originally published as:


DOI: http://doi.org/10.1785/0220150278
Seismological Research Letters

This copy is for distribution only by the authors of the article and their institutions in accordance with the Open Access Policy of the Seismological Society of America.

For more information see the publications section of the SSA website at www.seismosoc.org
The Engineering Strong-Motion Database: A Platform to Access Pan-European Accelerometric Data

by Lucia Luzzi, Rodolfo Puglia, Emiliano Russo, Maria D’Amico, Chiara Felicetta, Francesca Pacor, Giovanni Lanzano, Ulubey Çeken, John Clinton, Giovanni Costa, Llambro Duni, Esmael Farzaneh, Philippe Gueguen, Constantin Ionescu, Ioannis Kalogeras, Haluk Özener, Damiano Pesaresi, Reinoud Sleeman, Angelo Strollo, and Mehdi Zare

ABSTRACT

This article describes the Engineering Strong-Motion Database (ESM), developed in the framework of the European project Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation (NERA, see Data and Resources). ESM is specifically designed to provide end users only with quality-checked, uniformly processed strong-motion data and relevant parameters and has done so since 1969 in the Euro-Mediterranean region. The database was designed for a large variety of stakeholders (expert seismologists, earthquake engineers, students, and professionals) with a user-friendly and straightforward web interface.

Users can access earthquake and station information and download waveforms of events with magnitude \( \geq 4.0 \) (unprocessed and processed acceleration, velocity, and displacement, and acceleration and displacement response spectra at 5% damping). Specific tools are also available to users to process strong-motion data and select ground-motion suites for code-based seismic structural analyses.

INTRODUCTION

The repeated attempts of building unified engineering strong-motion databases in Europe are motivated by the increasing demand for strong-motion data, which are one of the primary sources of information used by engineering seismologists and earthquake engineers to predict ground shaking and to perform structural seismic analysis. Figure 1 summarizes the evolution of strong-motion data archiving in Europe since 1998.


During the 6th Framework Programme of the European Commission, within the Network of Research Infrastructures for European Seismology (NERIES) project (2006–2010, see Data and Resources), a new collection of accelerometric waveforms recorded in the Euro-Mediterranean region was promoted. At that time, only digital data were selected and 8000 waveforms, recorded in the time span 1995–2009 and in the 1.0–7.4 magnitude range, were gathered and uniformly processed with a fixed cutoff frequency of 0.1 Hz (Roca et al., 2011). Data were obtained mainly from Swiss, French, and Spanish collections, with the addition of sparse digital waveforms from Italian, Turkish, or Greek databases. Unfortunately, data are no longer available on the Web.

An additional strong-motion data collection was promoted within the Seismic Harmonization in Europe (SHARE) project (7th Framework Programme of the European Commission, 2010–2014) with the goal of testing the performance of candidate ground-motion prediction equations for the probabilistic seismic-hazard map of Europe (Yenier et al., 2010). Accelerograms were gathered from European or regional databases (e.g., the National Strong-Motion Network of Turkey [TR-NSMN], described in Akkar et al., 2010; Italian ACCELERometric Archive [ITACA] v.1.0, described in Luzzi et al., 2008 and Pacor, Paolucci, Luzzi, et al., 2011) as well as from worldwide databases (e.g., K-Net and KiK-net, see Data and Resources; the Next Generation Attenuation [NGA] database, Chiu et al., 2008; and a dataset compiled by Cauzzi and Facio, 2008). Neither metadata update nor uniform processing has been applied to waveforms (Yenier et al., 2010). This database contains 14,193 records from 2448 events in the 3.0–8.0 magnitude range, and only a flat file with strong-motion parameters is distributed at European Facilities for Earthquake Hazard & Risk (EFER; see Data and Resources).

A recent attempt to gather European strong-motion data was carried out within the Seismic Ground Motion Assessment (SIGMA) project (see Data and Resources). The Reference Database for Seismic Ground-Motion in Europe (RESORCE) database (Akkar et al., 2013), compiled within the Seismic Ground-Motion Assessment (SIGMA) project, contains pan-European strong-motion data and has the goal of improving seismic-hazard assessment in France and neighboring countries. The RESORCE database includes 5882 multicomponent and uniformly processed accelerograms from 1814 events in the

doi: 10.1785/0220150278
2.8 ≤ M_w ≤ 7.8 magnitude range, mainly obtained from the pan-European subset of the SHARE strong-motion collection. A set of 262 weak-motion waveforms, not included in the ESD, has been added for the hazard assessment of stable continental regions (e.g., France). Uniform processing has been applied (Akkar et al., 2013) to the raw data retrieved from various databases, although 89 already-processed multicomponent accelerograms from ESD have been directly incorporated because of the missing raw waveforms. To date, data are not public, although the database can be accessed upon request.

In contrast to the previous attempts, which were essentially project based, the approach followed within the Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation (NERA) project (2010–2014, Seventh Framework Programme of the European Commission; see Data and Resources) consisted of networking strong-motion data operators with the goal of creating a long-term infrastructure. The Engineering Strong-Motion Database (ESM, see Data and Resources) is specifically tailored to serve engineers and scientists alike in the assessment of seismic hazard, because it provides end users with quality-checked, uniformly processed strong-motion data from events in the 4.0–7.5 magnitude range (where the minimum magnitude is selected according to the lower threshold generally employed for engineering applications and hazard assessment). If magnitudes larger than 4 are considered, ESM encompasses all the previous European databases and largely overtakes all of them, because it is closely linked to the European Integrated Data Archive (EIDA) (Fig. 1), a key infrastructure aimed at archiving digital waveforms that will be discussed in the following sections.

The ESM database is complementary to the Rapid Response Strong-Motion Database (RRSM, see Data and Resources and Fig. 1), also developed in the framework of Project NERA (see Cauzzi et al., 2016). RRSM allows rapid access to earthquake information, peak-ground-motion parameters, and response spectral amplitudes within minutes after an earthquake with magnitude ≥ 3.5. Data are exclusively obtained from EIDA and are automatically processed and published on the Web without manual checks. The complete data collection dates back to 2005.

This article describes the data collection strategy adopted for ESM, the content of the database, the way of accessing and disseminating data, the tools supplied to end users and the long-term strategy for database maintenance and update.

**ESM DATA COLLECTION STRATEGY**

Examining the state of the art of strong-motion data collection in Europe (Fig. 1), the understanding is that too many transitions were made after the initial database attempt (ESD). Therefore, before undertaking the construction of the ESM database, a step back was made to retrieve the unprocessed data (i.e., not available in the EIDA platform) from the original databases. Toward this goal, three main regional databases were identified:

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

Finally, pan-European data from 1972 to 2008, not included in regional databases, are extracted from the ESD (see Data and Resources).

Raw waveforms extracted from these databases form the initial core of the ESM data collection and represent the following percentages of the current dataset: HEAD ~2%, ITACA ~20%, TR-NSMN ~7%, and ESD ~1%.

A common characteristic of “historical” strong-motion data is the lack of standardized metadata and formats for their full integration with actual seismological standards. Therefore, a major effort has been spent toward the homogenization and standardization of metadata.

Standardized network codes, assigned by the International Federation of Digital Seismograph Networks (FDSN, see Data and Resources), are assigned to all providers, including the ones no longer operating (e.g., networks belonging to the former Yugoslavia), whereas waveforms are named following the Standard for Exchange of Earthquake Data (SEED) convention (see Data and Resources).

Nevertheless, the innovation of ESM is not only the preservation of the existing patrimony of historical strong-motion data and their harmonization with actual seismological standards. A major innovation is the full exploitation of the potential of the new generation data centers.

Currently, new generation instruments can record weak-to-strong motions and transmit waveforms in real time to data centers, soon after the occurrence of a seismic event. A recent initiative within Observatories and Research Facilities for Euro-
European Seismology (ORFEUS, see Data and Resources) is EIDA (see Data and Resources), structured as a distributed data center to securely archive seismic waveforms gathered by European research infrastructures and to provide transparent access to the archives. Several European data centers act as EIDA nodes, collecting and archiving data from seismic networks and deploying broadband and short-period sensors, accelerometers, infrasound sensors, or other geophysical instruments. The Incorporated Research Institutions for Seismology (IRIS) Data Center also supports access to time-series data, related metadata, and event parameters. Both ORFEUS and IRIS are coordinated within the International Federation of Digital Seismograph Networks (FDSN, see Data and Resources). The availability of continuous data streams allows access to seismic signals in quasi-real time and to progressively populate the ESM database through semiautomatic procedures. In the following, the procedure for retrieving and publishing strong-motion data into ESM after an earthquake occurrence in the pan-European region is described.

Step 1: Data Upload after an Event Alert
After the occurrence of any event with magnitude larger than or equal to 4, reported by the European–Mediterranean Seismological Centre, an automatic procedure for signal windowing is applied to the continuous streams available through EIDA and/or IRIS data centers (see Data and Resources), and waveforms are uploaded into ESM together with their metadata. At this stage, only preliminary location parameters are assigned to the earthquake. Offline waveforms, for example, data not available through webservies, are then uploaded after the network operators make them available.

Step 2: Data Processing and Publishing
Before the manual revision, waveforms are automatically processed using the procedure proposed by Paolucci et al. (2011), described in detail in the Waveform Processing section. Automatically processed waveforms are stored in the database, although not accessible to users, and only the peak ground acceleration (PGA) is published. Subsequently, waveforms are manually revised and processed, and acceleration, velocity, and displacement time series are released, together with acceleration and displacement response spectra at 5% damping. Bad-quality records (e.g., noisy records or waveforms containing spikes) are made available to users only in the unprocessed version. The release time of processed waveforms ranges from a few hours (relevant events) to a few days.

Step 3: Metadata Revision
Event and station metadata contained in the database are periodically revised. The event information collected from earthquake-specific literature studies are always ranked as the primary reference for large seismic events. For moderate to small events, the sources of information are regional catalogs (e.g., the Istituto Nazionale di Geofisica e Vulcanologia Bulletin, see Data and Resources) or the Bulletin of the International Seismological Centre (ISC, see Data and Resources), in case regional catalogs are unavailable. The ISC bulletin relies on contributions from worldwide seismological agencies and is typically 24 months behind real time. Different magnitudes (e.g., $M_w, M_l, M_b, M_s$) are reported in the database, as well as moment tensor solutions from different agencies. Information on the geometries of the seismic sources comes from regional or international catalogs (e.g., Ambraseys, Douglas, et al., 2004, for several European events, Database of Individual Seismogenic Sources [DISS] for Italy; Greek Database of Seismogenic Sources [GREDASS] for Greece, and finite-source rupture model database [SRCMOD] for large events that occurred worldwide; see Data and Resources) or from specific source-model studies.

Station metadata are periodically updated after specific studies are published in the literature or after the results of national and international projects. The station information actually contained in the ESM is obtained from regional databases (ITACA, TR-NSMN, and HEAD; see Data and Resources) or from specific literature studies (e.g., Zare et al., 1999; Régnier et al., 2010; Michel et al., 2014).

DESCRIPTION OF DATA CONTAINED IN ESM
The ESM database contains 23,000 three-component waveforms. About 60% of them (13,191) are manually processed, 15% are automatically processed and need manual revision, and 25% are judged of bad quality. Bad-quality data are preserved in the database, because intensity measures such as PGA can be used with good confidence. The manually processed waveforms derive from 1929 seismic events ($M_w \geq 4.0$) and were recorded by 1901 sites operated by 38 networks. The following statistics are relative to this subset of waveforms.

Figure 2 shows the magnitude–distance sampling of the records, whereas in Figure 3 the distributions of event depths, style of faulting, and maximum usable period are shown in terms of the number of records. The ESM dataset is well sampled in the 4.0–6.0 magnitude range and in the epicentral distance range 10–200 km (Fig. 2), with a significant number of waveforms related to strong events ($6.0 < M_w < 7.5$) recorded at epicentral distances larger than 10 km. The strongest events have been recorded in Italy (1980 Irpinia, $M_w$ 6.9), Turkey (e.g., 1999 İzmit, $M_w$ 7.4; 1999 Düzce, $M_w$ 7.1), and in Iran (1990 western Iran, $M_w$ 7.4). As shown in Figure 3, most of the events in the database are shallow crustal earthquakes with hypocentral depth lower than 40 km. The majority of records are related to normal (~30%) and reverse (~25%) faulting, although a considerable number of records are still not associated with a style of faulting (~30%). The maximum usable period for the horizontal and vertical components is also shown in Figure 3c,d. The majority of waveforms have a maximum usable period ranging from 2.5 to 10 s. About 1000 waveforms have lower values (1.0–2.5 s), mainly because they are recorded by analog instruments, and only a limited number of records (~500) can be used up to 20 s.

Several recording stations are characterized by geotechnical and geophysical measurements (available for several Albanian, French, Iranian, Italian, Swiss, and Turkish stations). Geophysical measurements, in particular, consist of about 340 1D velocity
profiles estimated through different geophysical prospection methods, such as (1) invasive active (crosshole and downhole), (2) noninvasive passive (e.g., microtremor array measurements), and (3) noninvasive active (e.g., multichannel analysis of the surface waves). The average shear-wave velocity of the uppermost 30 m ($V_{s30}$), a fundamental information for engineering applications, has been calculated from the available velocity profile or, in a few cases, is taken from the literature. Figure 4a shows that the majority of stations have $V_{s30}$ in the range 200–600 m/s, whereas very few stations are characterized by $V_{s30} > 800$ m/s (rock conditions). Because the percentage of sites associated with a $V_{s30}$ value is quite low (18%), in order to characterize a larger number of sites, we make use of the support of surface geology. Figure 4b shows the distribution of the Eurocode 8 (EC8) subsoil categories, evaluated from the measured $V_{s30}$ or inferred from surface geology. Despite the support of surface geology, information for a remarkable number of sites (32%) is still missing.

**DATA ACCESS AND DISSEMINATION**

The ESM website is organized in three main blocks, relevant to waveforms, recording stations, and seismic events. Seismic events can be retrieved entering the “Events” page of the portal. The user can select 13 parameters, including date and time of the event, magnitude range, hypocentral coordinates, or style of faulting. The query returns a list of earthquakes that can be individually accessed.

Figure 5 shows the example of the event occurred on 7 April 2014 at 19:27:01 UTC ($M_w$ 4.9) at the border between Italy and France. The event was recorded by 105 stations belonging to seven different networks (codes IT, RA, CH, GU, G, FR, and MN) and is associated with the identities of major international catalogs to provide complete information to the user. In this example, the preferred location is by the Helmholtz-Centre Potsdam—GFZ German Research Centre for Geosciences (see Data and Resources), whereas several magnitude determinations are attributed from international agencies. Focal mechanism solutions and style of faulting are also provided. The list of stations that recorded the event is reported at the bottom of the page and includes metadata such as EC8 site class, epicentral distance, and the maximum peak ground motion (e.g., PGA, peak ground velocity, and peak ground displacement) of the three components.

Station information can be accessed entering the “Stations” page, where recording sites can be retrieved according to 14 parameters, including location, network and station code, and parameters related to the site characterization, such as the average velocity in the uppermost 30 m ($V_{s30}$). Figure 6 shows the example for the station AQV, belonging to the network IT (Italian accelerometric network, operated by the Italian Department of Civil Protection). The station is displayed on a topographic map, and information related to location, housing, and site class are provided. Metadata related to each station are also available in the form of a report (accessed via the “Monography” button on the station page) containing detailed information, such as stratigraphic and geophysical logs or the horizontal-to-vertical spectral ratio obtained from noise measurements.

Waveform information can be accessed entering the “Waveforms” page, where 35 parameters can be specified related to stations, events, or waveform metadata. Waveforms can be explored with the aid of a visualization tool that allows zooming and exporting the time series as images.

Upon user registration, time series can be downloaded in ASCII format, as unprocessed acceleration time series, as processed acceleration, velocity, and displacement time series, or as acceleration, pseudovelocity, and displacement response spectra (5% damping) calculated at 105 periods (0.01–10 s). A client, written in Python language, can be downloaded from the ESM homepage to convert ASCII files in standard seismological formats (e.g., SAC or miniSEED).

**Data Citation, Acknowledgments, and License for Data Distribution**

Each waveform can be tracked, because the metadata that reproduce the complete path, from the original data source to the processed data, are included. The appropriate citation is also reported, together with the network digital object identifier (DOI), when available. A license can be provided by network operators, following the Creative Commons standards (see Data and Resources), to enable the sharing and use of data through free legal tools and guarantee visibility to the original author. Figure 7 shows an example of citation and acknowledgment for the E component of a waveform recorded by the station MLR, operated by Institutul National de Cercetare si Dezvoltare pentru Fizica Pamantului (INFP) (National Institute for Earth Physics of Bucharest, network code RO), relative to the event that occurred on 30 January 2008 at 13:09:30 (UTC).
ADDITIONAL ESM TOOLS

Waveform Processing
A waveform processing web front end is available at the ESM website (see Data and Resources), providing access to all waveforms included in the ESM database. Individual acceleration time series are processed manually following the general procedure described in Paolucci et al. (2011), which consists of

- linear detrending of the uncorrected acceleration signal (subtraction of a first-order polynomial);
- application of a cosine taper, at the beginning and end of the signal, with percentage fixed to 5% of the signal length and with the possibility of being modified by the user;
- visual inspection of the Fourier spectrum to select the band-pass frequency range (band-pass frequency may be different for the three components);
- application of a second-order acausal time-domain Butterworth filter to the acceleration time series, with zero-pads added at the beginning and end of the signal before the acausal filter is applied (Boore, 2005);
- removal of zero-pads from the acceleration trace;
- (begin/end) taper of the acceleration signal, with percentage fixed to 5%;
- computation of the velocity signal and linear detrend;
- (begin/end) taper of the velocity signal, with percentage fixed to 5%;
- computation of displacement signal and linear detrend;
- (begin/end) taper of the displacement signal, with percentage fixed to 5%; and
- recursive differentiation to obtain the velocity and the acceleration time series, respectively.

Acausal filters are preferred to causal, because elastic and inelastic response spectra are sensitive to the corner periods used in causal filtering at periods much shorter than the corner periods (Boore and Akkar, 2003).

Figure 8 shows two screenshots representing the processing web front end of the three components of the ground motion recorded at station NIR (Iran) on 28 February 1997 at 12:57:22 Greenwich Mean Time (GMT) \( M_w 6.0 \). Figure 8a displays the three components of the unprocessed and processed acceleration time series, whereas Figure 8b displays the corresponding Fourier spectra, which are useful to check the effect of the filtering. The following parameters can be modified by the user:

- length of the waveform cut (begin/end, in seconds);
- percentage of signal tapering;
- flag normal/late-triggered (in which “late-triggered” is a signal triggered by \( S \) waves or later phases), which implies
different processing (details are in Pacor, Paolucci, Ameri, et al., 2011, and Paolucci et al., 2011):
- output sampling interval;
- constant for the multiplication of the signal; and
- band-pass corners (Hz) and order of the acausal Butterworth filter.
Registered users can select waveforms, perform customized processing, and save the results.
Spectrum-Compatible Data Selection

The ESM database is coupled with the REXELite application, which is the online version of the computer program REXEL (Iervolino et al., 2009, 2011), for the selection of ground-motion suites for code-based seismic structural analyses. REXELite allows searching for combinations of seven 1- or 2-component strong-motion records, compatible, in average, with a specified code spectrum. More specifically, REXELite (1) automatically builds code spectra for any limit state according to Eurocode 8 (2002; EC8) and (2) finds the set of seven records having the most similar spectral shape with respect to that of the code, and for which the average also matches the target spectrum, in a user-specified period range and with a desired tolerance. The records are pre-selected by the user according to specific features, such as magnitude and source-to-site distance ranges, style of faulting, and soil conditions, codified as EC8 site classes. The resulting set of accelerograms may include unscaled (original) or amplitude-scaled records and may be used for code-compliant nonlinear time history analyses of structures.

DISCUSSION

We described the ESM developed in the framework of the European project NERA. ESM is tailored to enable users to fully exploit pan-European strong-motion data recorded since 1969 relative to events with magnitude larger than or equal to 4. The database has been designed for a large variety of stakeholders (expert seismologists, earthquake engineers, students, and professionals); and, for this reason, the web interface is friendly and straightforward. In addition, expert users may benefit from specific tools for data processing and data selection.

The core of ESM was built from existing regional databases (~30% of the actual waveforms) and is constantly growing, thanks to the continuous supply of waveforms gathered from EIDA or offline archives by several European providers. The rate of growth of the database is about 3000 waveforms per year, if we exclude seismic sequences that could double the estimated rate.

ESM is not only the result of a project but is also part of a long-term vision for the distribution of strong-motion data in
Figure 7. (a) Waveform recorded at station MLR on 30 January 2008 at 13:09:30 (Greenwich mean time [GMT]) by Institutul National de Cercetare si Dezvoltare pentru Fizica Pamantului (INFP) (network code RO) and (b) with the licensing and acknowledgments section, with citations of the waveform. The color version of this figure is available only in the electronic edition.

Figure 8. Three components of the ground motion recorded at NIR (Iran) on 28 February 1997 at 12:57:22 GMT (MW 6.0): (a) superimposed unprocessed and processed acceleration time series; (b) Fourier spectra of unprocessed and processed acceleration time series. The color version of this figure is available only in the electronic edition.
Europe. A thematic working group has been established within ORFEUS (WG5—acceleration and strong-motion data) to create a network of strong-motion data operators in Europe. They are directly involved in the decisional process (e.g., setting rules for data dissemination) and are continuously updated on the technological progress and on the state of the art of techniques for metadata compilation and data processing. To ensure infrastructure sustainability, ORFEUS will foster the usage of standard FDSN webservises and promote the exposure of data through the EIDA-distributed archive.

Data should be not only distributed and exchanged in an open data framework but also interoperable with other disciplines. Toward this goal, the ESM has been selected as one of the infrastructures of the European Plate Observing System (EPOS, see Data and Resources), a long-term plan for the integration of national and transnational research infrastructures for solid Earth science in Europe, to provide seamless access to data, services, and facilities.

DATA AND RESOURCES


ACKNOWLEDGMENTS

The Engineering Strong-Motion Database (ESM) interface was designed and implemented by Istituto Nazionale di Geofisica e Vulcanologia (INGV) using funding from the European Commission-funded Seventh Framework Programme in the Network of European Research Infrastructures for Earthquake Risk Assessment and Mitigation (NERA) project (see Data and Resources). We are grateful to the NERA-NA3 working group (coordinated by Sinan Akkar) for supporting the development of this work. We thank the European Integrated Data Archive (EIDA) and Observatories and Research Facilities for European Seismology (ORFEUS) Working Group 5 (acceleration and strong-motion data) communities for providing open data to the community and for the fruitful discussions on data policy.

We acknowledge the following institutions and researchers for providing local databases, original waveforms, or metadata or for being part of the ORFEUS-WG5: Laboratory of Institute of Engineering Seismology and Earthquake Engineering, Thessaloniki (ITSAK), Greece, which provided the Unified HEllenic Accelerogram Database (HEAD) data (1973–2000); P. Dimitriou (deceased); B. Margaris; C. Papaioannou; C. Papazachos, A. Skarlatooudis, N. Theodoulidis (ORFEUS W5G member and also for his contribution in the discussions during the NERA meetings); National Observatory of Athens (NOA), Greece: HEAD; Italian Department of Civil Protection (DPC) and Istituto Nazionale di Geofisica e Vulcanologia (INGV): the ITalian ACcelerometric Archive (ITACA); Disaster and Emergency Management Presidency of Turkey (AFAD); National Strong-Motion Network of Turkey (TR-NSMN) from 1976 to 2007; Earthquake Engineering Research Centre (EERC) of the University of Iceland: Internet site for European Strong Motion Database (ESD); Building and Housing Research Center (BHRC), Iran; Institute of GeoSciences, Energy, Water, and Environment (IGEWE), Tirana, Albania; Helmholtz Centre Potsdam (GFZ) German Research Centre for Geosciences; French seismologic and geodetic network (RESIF); Bogazici University Kandilli Observatory and Earthquake Research Institute (KOERI), Turkey; National Institute for Earth Physics (NIEP), Bucharest, Romania; Swiss Seismological Service (SED); Friuli accelerometric network

REFERENCES


