2012 EMILIA EARTHQUAKES

AIDA – Seismic data acquisition, processing, storage and distribution at the National Earthquake Center, INGV

Salvatore Mazza^{*}, Alberto Basili, Andrea Bono, Valentino Lauciani, Alfonso Giovanni Mandiello, Carlo Marocchi, Francesco Mele, Stefano Pintore, Matteo Quintiliani, Laura Scognamiglio, Giulio Selvaggi

Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Roma, Italy

Article history

Received July 24, 2012; accepted September 10, 2012.

Subject classification:

Surveys, measurements and monitoring, Earthquake location, Moment tensor, Magnitude.

1. Introduction

On May 4, 2012, a new system, known as the AIDA (Advanced Information and Data Acquisition) system for seismology, became operational as the primary tool to monitor, analyze, store and distribute seismograms from the Italian National Seismic Network. Only 16 days later, on May 20, 2012, northern Italy was struck by a M_L 5.9 earthquake that caused seven casualties. This was followed by numerous small to moderate earthquakes, with some over M_L 5. Then, on May 29, 2012, a M_L 5.8 earthquake resulted in 17 more victims and left about 14,000 people homeless. This sequence produced more than 2,100 events over 40 days, and it was still active at the end of June 2012, with minor earthquakes at a rate of about 20 events per day.

The new AIDA data management system was designed and implemented, among other things, to exploit the recent huge upgrade of the Italian Seismic Network (in terms of the number and quality of stations) and to overcome the limitations of the previous system. Its major achievements are:

- real-time data acquisition from the Istituto Nazionale di Geofisica e Vulcanologia (INGV; National Institute of Geophysics and Volcanology) seismic network, and other external networks;
- real-time continuous data archiving in standard formats;
- data dissemination for scientific purposes;
- near-real-time automatic earthquake detection and hypocenter calculation (e.g., moment tensors, shake-maps);
- database archiving of all parametric results;
- close interactions with existing procedures within the INGV seismic monitoring environment;
- high availability and robustness through hardware and software redundancy;
- automatic or semi-automatic configuration routines.

2. Seismic data acquisition

As shown in Figure 1, the core of Centro Nazionale Terremoti (CNT; National Earthquake Center) acquisition system is based on the SeedLink protocol, a de-facto stan-

dard originally created as a transport layer for SeisComP (<http://www.seiscomp3.org/>). It includes the tools to collect, distribute and archive waveform data, and it has proven to be a reliable and robust product. Stations are connected through many different links, but with only two protocols: SeedLink, in use with all of the stations equipped with SeedLink locally (e.g., Quanterra or INGV home-developed GAIA digitizers) or with the SeedLink server of other network operators (co-operating Institutes or other INGV sections); and the Nanometrics protocol, which is adopted for stations equipped with Nanometrics instrumentation, as either transmitting with the Nanometrics Lybra system or provided by partners adopting the protocol. A plugin *nmxptool* [Quintiliani 2007] is available to get data from the Nanometrics servers in near-real-time to SeedLink, and to archive them in miniSEED format.

Indeed, the SeedLink server acts as a concentrator, and also as an integrator, which allows prompt use of data in real-time for analysis, distribution and archiving. Data in real-time are distributed only to INGV partner Institutes (national and international), as a bilateral exchange of data or on the basis of official agreements.

3. Data archiving and distribution

The CNT seismic archive holds 22.3 Tbytes of data at present, which is growing at a rate of about 5 Tbytes/year. These data are collected from 22 networks, for a total of 422 stations, 330 of which are managed by INGV directly, with the others as contributions from local or foreign networks. Most of the stations are equipped with broadband velocimeters, which are complemented by accelerometers in about 120 stations (see Figure 2 for a map of the integrated National Network; i.e., including stations from all of the contributing institutes). In Appendix A, the archived data statistics are provided as data amounts per year and as data contributions to the archive per network.

The MedNet network archived data go back to 1990, while

Figure 1. Scheme of the AIDA system demonstrating the data flow from stations (SeedLink and Nanometrics protocols) to real-time analysis and archiving. ADS, Seismic Data Analysis; CNT, National Earthquake Center; DB, database; SAN, Storage Area Network; TDMT, time-domain moment tensors.

Figure 2. Map of the extended national network on which the seismic survey of Italy is founded. The stations are contributed by national and international institutes, and are mostly equipped with high-resolution broadband instrumentation.

the National Network data are available only from 2007, when the new archiving system was started. The National Network data before 2007 are not included yet, although their inclusion has been planned. They are, however, mostly short-period event data. In general, the data are freely available, although those produced by specific projects and experiments can be restricted for an agreed period of time (typically 1-2 years, after which time they must be released). We have adopted the approach proposed in and realized by the European Network of Research Infrastructures for European Seismology (NERIES) project (<http://www.neries-eu.org/>), of a distributed archive: the European Integrated Data Archive (EIDA). This approach results in a highly scalable archive, in which a strong backbone is provided by the major institutes, but where minor participants can also have their role. Each data center can manage and look after their data and station information from its privileged workplace. Users can access the same integrated dataset through contacting one of the participating data centers, without knowing where the data are actually stored. In general, the CNT Data Center archives data collected from the partner institutes for its internal purposes only (e.g., reproducibility of location and event parameters), and does not distribute these unless an explicit agreement exists to do so.

Data archiving is initially carried out in real-time into a SEED data structure, and the data are made available to the community as soon as they are collected (usually with a few seconds delay). However, the data archiving does not end

Figure 3. Above: (a) Comparison between synthetic and recorded traces (six stations of 64 used) for the moment tensor automatic solution of the Emilia May 20, 2012, M_W 5.9 earthquake. Below: (b) Map of the stations that contributed to the moment tensor automatic solution for the Emilia May 20, 2012, earthquake.

here. The day after their collection, all of the data are moved to a Storage Area Network and quality checked. In case of gaps due to link failures or station communication malfunction, the data are recovered from the stations (or from the servers) as soon as the link is re-established.

To recover data gaps, we use two specially developed procedures daily: Offline Archive Completion gets the waveform data from the Nanometrics Data Server buffer, which holds the station data from ca. 20 days; Mini Seed Data Completion recovers data gaps directly from the INGV stations, where the SeedLink servers store the data in a local archive (Quanterra and GAIA digitizers). It is important to realize that data extracted from these archives within a few minutes after an earthquake can be significantly improved by off-line data recovery in the following days.

Relevant efforts have been devoted to establish protocols for exchanging station information between the network co-operators and the Seismic Data Center, with the excellent result of very reliable station information, and consequently robust seismic parameter estimates.

4. Real-time earthquake analysis and quasi-real-time revision

Real-time earthquake evaluation consists of automatic location, and local magnitude and moment tensor estimations. The AIDA real-time analysis relies on two main components: the Earthworm system [Johnson et al. 1995], which is used to

automatically locate events and to estimate local magnitudes; and a set of databases specifically developed to store, use and distribute the seismic parameters in real-time. The particular shape of Italy, as surrounded by about 8,000 km of coast, and the heterogeneity of its crust, which can range from stable continental to recent oceanic within a few tens of kilometers, ren-

der the task of locating earthquakes in Italy quite challenging. To face the difficulties due to the structure of the country and to the inhomogeneous station distribution, we have adopted the Earthworm software, because of its high configurability.

To deal with the large number of parameters of the Earthworm system, and also with station instrumental heterogeneity, we have developed ad-hoc procedures to automatically generate configurations from the centralized MySQL database SeisNet, where all of the information regarding the servers and stations is stored. Similarly, we store all types of Earthworm messages in the main CNT database for earthquake parameters (SeisEv) by means of the Mole system [Quintiliani and Pintore 2012; software available at <http://earthworm.isti.com/trac/earthworm/browser/trunk/src/archiving/mole>]. Although the real-time algorithms for the detection and association of seismic phases are new, the final earthquake parameters are the result of an interactive analysis that has remained unchanged (the automatic performance of the Earthworm system in Italy is described below). All of the automatic estimations are evaluated and revised by the INGV Seismic Service personnel before they are communicated to the Civil Protection Agency and then published through email, SMS, and web pages. In order to let seismologists move seamlessly from the previous system to the new one, we modified the traditional interactive software for event revision that had been in use since the early development of the Italian Digital Seismic Network [Amato et al. 2006] to access the SeisEv and the SeisNet databases. The algorithms to re-locate earthquakes and re-compute their local magnitudes have not been modified. Likewise, we have adopted the same crustal velocity model and attenuation function that has been in use since April 2005 [Mele et al. 2010], to ensure compatibility with the Italian Seismic Bulletin.

Following earthquake parametric determination by Earthworm, an independent procedure is automatically initiated to compute moment tensors with the time-domain technique (TDMT) [Scognamiglio et al. 2009, Dreger and Helmberger 1993], which has been improved by an automated algorithm for station selection that optimizes distance and azimuthal coverage. After the automatic computation of a TDMT solution, the seismologists can revise the mechanism with a computer-assisted procedure aimed at retaining or rejecting single stations, while the algorithm chooses the optimal coverage among the others. As an example, Figure 3a shows the automatic solution for the Emilia, northern Italy, main shock (May 20, 2012), along with the excellent fit between the synthetics and the recordings for six stations (out of the 64 used; Figure 3B) [see also Scognamiglio et al. 2012, this volume].

5. Parametric and waveform data dissemination

Parametric data are published on the CNT website (<http://cnt.rm.ingv.it>) and the Italian Seismic Instrumental and parametric DatabasE (ISIDe) website (<http://iside.rm.ingv.it/>) in near real-time, immediately after the earthquake location

and magnitude have been revised. The CNT web pages show only the events already communicated to the Italian Civil Protection, with all of the relevant available information; i.e., historical and recent seismicity, ShakeMaps, phases, and so on. ISIDe provides the best revised information on very recent seismicity as soon as it is available; it includes both quasi-real-time revisions and the Italian Seismic Bulletin. The database today includes more than 88,000 earthquakes that have occurred in Italy and in surrounding seas since April 16, 2005, which have ranged from M_L 0.2 to M_L 5.9.

Although ISIDe covers only the last few years of Italian seismicity, it is unprecedented in Italy for its completeness and homogeneity, and it represents an optimum test set to verify small and moderate seismicity pattern models [Schorlemmer et al. 2010]. More than 4,900 new users registered on the ISIDe web portal immediately after the Emilia May 20 and 29, 2012, earthquakes, thus reaching a total of 10,000 users. There were ca. 4 million visitors to the CNT web pages during the first month of this sequence, and ca. 0.5 million to the ISIDe website.

Event data are available from ISIDe, as tar sac files. Continuous data are available from <http://eida.rm.ingv.it/> through an interactive form, which gives access to the National Seismic Network data and also to the data provided by the EIDA. Although user friendly, the interface is not suitable for massive data extraction. To overcome this limitation, we have implemented a web-services infrastructure for EIDA to share data efficiently and to provide a high security level. Skilled users are allowed to implement their own data retrievers. The INGV web services system (*ingv_ws_data*) is mainly composed of three tools:

- a database to store information about data requests and users;
- a service provider that handles requests and recovers data;
- a java client application to download data, which is downloadable from <http://webservices.rm.ingv.it>.

A total of about 400 Gbyte of data (9,950 requests) were downloaded during the Emilia seismic crisis from the May 16 to June 16, 2012, from the web services through the INGV client. Also, 220 Gbyte more (corresponding to 125,000 requests) were extracted via different tools (like the web interactive interface, or tools connected directly to the ArcLink server).

6. Considerations on real-time performance

We have investigated the AIDA real-time performance with particular regard to hypocenter locations and magnitude estimates. Although the system shows quite relevant hypocentral discrepancies from the true revised locations at low magnitudes (Figure 4a), only a few events with magnitudes >3.5 are affected by an evident error in the real-time locations. These errors are not surprising during a seismic sequence such as in Emilia, with two main events of M_L 5.9

Figure 4. (a) Distribution of the mislocations as a function of the revised magnitudes. (b) Distribution of the differences between the real-time and revised magnitudes. (c) Differences between the real-time and revised magnitudes as a function of the distances between the real-time and revised locations. (d) Distribution and cumulative distribution of the distances between the real-time and revised epicenters. (e) Distribution and cumulative distribution of the distances between the real-time and the revised magnitudes.

and 5.8, an activated thrust fault system ca. 50 km long, and thousands of aftershocks in a few days. Under such conditions, it is exceedingly difficult for most automatic procedures to discriminate events. At the same time, only a small number of events with magnitudes >3.5 showed notable inaccuracies in their real-time magnitudes (Figure 4b), which can be attributed to different causes. It is well known that unless the correct station corrections are applied, the superficial quaternary sediments present in the Po Plain can strongly influence seismic wave amplitudes and result in large magnitude overestimations. Also, mislocations can be identified as another source of differences between automatic and revised magnitudes, by observing the strong correlation between magnitude and location error (Figure 4c). This correlation also explains why positive errors in real-time magnitudes are predominant over negative errors: a large error in the location has the effect of erroneously augmenting the average hypocentral distance, thus causing larger distance corrections. Finally, inaccuracies originate from the human error that can occur when locations and magnitudes are calculated only a few minutes after the occurrence of the event by personnel on duty who are working under particular pressure.

As shown in Figure 4d, about 60% of the real-time epicenters are within 10 km of the revised ones, and about 80% within 50 km. The distribution of the real-time magnitudes (Figure 4e) is strongly asymmetric; about 60% of the magnitudes computed in real-time are within the revised magnitudes ± 0.2 ; only 2% of the total shows a difference <-0.2 , while the remaining 38% show a difference $>+0.2$, mainly because of mislocations.

In general, according to our experience, the difficulties encountered in earthquake localization and magnitude estimation are related more to the geomorphological settings of the Po Plain and to the uneven station distribution than to shortcomings of the Earthworm system. When the Emilia sequence began, the station distribution was so poor that the epicentral distance of the closest station was 17 km, and that of the second closest 30 km. Appendix B gives a map of the available stations in the area, as well as a diagram with station up-times. Although the installation of new stations in the epicentral area was very fast, the station coverage during the very first hours was not optimal.

7. Conclusions and developments

A few days before the seismic sequence started on May 20, 2012, in Emilia (northern Italy), the new AIDA system had just become operational as a comprehensive tool to monitor, analyze, store and distribute the Italian National Seismic Network seismograms. During this sequence, the system performance met expectations in terms of service continuity and the robustness of real-time acquisition, automatic earthquake detection, and dissemination of wave-

forms and event parameters.

However, this recent crisis has made it clear that greater efforts are needed to improve location and magnitude estimations in the area. Location accuracy can be increased by adopting a three-dimensional model (or station velocity profiles). For magnitude calculations, we have adopted the same Hutton and Boore [1987] relationship that was in use in the old system, to ensure continuity in magnitude evaluation. It was evident during the Emilia sequence that some adjustments to these Hutton and Boore [1987] parameters are necessary to account for the crustal characteristics of Italy [e.g., Gasperini 2002]. The application of a poor attenuation relation is the cause of overestimations for magnitudes >3.5 . Station corrections must be introduced [see Scognamiglio et al. 2012] for the stations where the recordings are influenced by strong site effects due to the thick sediments in the Po Plain.

References

- Amato, A., L. Badiali, M. Cattaneo, A. Delladio, F. Doumaz and F. Mele (2006). The real-time earthquake monitoring system in Italy, *Géosciences Révue du BRGM, Les risques telluriques*, 6 (September 2006); <http://www.brgm.fr/dcenewsFile?ID=294>
- Dreger, D.S., and D.V. Helmberger (1993). Determination of source parameters at regional distances with three-component sparse network data, *J. Geophys. Res.*, 98, 8107-8125.
- Gasperini, P. (2002). Local magnitude revaluation for recent Italian earthquakes (1981-1996), *J. Seismol.*, 6, 503-524.
- Hutton, L.K., and M. Boore (1987). The M_L scale in southern California, *B. Seismol. Soc. Am.*, 77, 2074-2094.
- Johnson, C.E., A. Bittenbinder, B. Bogaert, L. Dietz and W. Kohler (1995). Earthworm: a flexible approach to seismic network processing, *IRIS Newsletter*, 14, 1-4.
- Mele, F., L. Arcoraci, P. Battelli, M. Berardi, C. Castellano, G. Lozzi, A. Marchetti, A. Nardi, M. Pirro and A. Rossi (2010). *Bollettino Sismico Italiano 2008, Quaderni di Geofisica e Vulcanologia*, 85 (in Italian); <http://hdl.handle.net/2122/6159>
- Quintiliani, M. (2007). libnmxp e nmxttool: software Open-Source per trasmissioni dati sismici Nanometrics, *Rapporti Tecnici INGV*, 51, 25 pp.; <http://istituto.ingv.it/l-ingv/produzione-scientifica/rapporti-tecnici-ingv/archivio>
- Quintiliani, M., and S. Pintore (2012). Mole: an open near-real-time database-centric Earthworm subsystem, *Seismol. Res. Lett.*, submitted.
- Schorlemmer, D., F. Mele and W. Marzocchi (2010). A completeness analysis of the national seismic network of Italy, *J. Geophys. Res.*, 115, B04308; doi:10.1029/2008JB006097.
- Scognamiglio, L., E. Tinti and A. Michelini (2009). Real-time determination of seismic moment tensor for the Italian region, *B. Seismol. Soc. Am.*, 99, 2223-2342.

Scognamiglio, L., L. Margheriti, F.M. Mele, E. Tinti, A. Bono, P. De Gori, V. Lauciani, F.P. Lucente, A.G. Mandiello, C. Marcocci, S. Mazza, S. Pintore and M. Quintiliani (2012). The 2012 Pianura Padana Emiliana seismic sequence: locations, moment tensors and magnitudes, *Annals of Geophysics*, 55 (4); doi:10.4401/ag-6159.

*Corresponding author: Salvatore Mazza

Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Roma, Italy; email: salvatore.mazza@ingv.it.

© 2012 by the Istituto Nazionale di Geofisica e Vulcanologia. All rights reserved.

Appendix A

Figure A1. Amount of data archived in the CNT, INGV, database since 1990 (MedNet) to the present (MedNet + INGV National Network + other collaborating networks).

Figure A2. Contributions of other Italian and European seismic networks to the CNT archive.

Appendix B

Figure B1. Map of the permanent (green) and portable (yellow) seismic stations in the area of the Emilia seismic sequence.

Figure B2. Up-time diagram of the stations in the epicentral area.