

# INGV data lifecycle management system performances during Mw 6.0 2016 Amatrice Earthquake Sequence

STEFANO PINTORE, FABRIZIO BERNARDI, ANDREA BONO, PETER DANECEK, LICIA FAENZA, MASSIMO FARES, VALENTINO LAUCIANI, FRANCESCO PIO LUCENTE, CARLO MARCOCCI, DONATELLA PIETRANGELI, MATTEO QUINTILIANI, SALVATORE MAZZA, ALBERTO MICHELINI

Istituto Nazionale di Geofisica e Vulcanologia

stefano.pintore@ingv.it

## Abstract

*At 01:36:32 UTC on August 24, 2016 an earthquake of magnitude 6.0 occurred in Central Italy, affecting many small towns and municipalities in the Lazio, Umbria, Marche and Abruzzo regions. The event caused severe damages, many victims and 299 fatalities. Only 21 seconds after the beginning of the earthquake, the first automatic location of this earthquake was available and stored in our earthquakes database. The first magnitude estimate followed 68 seconds after the origin time. Few seconds later the INGV seismologists on duty in accordance to the agreed protocols provided the first alert to the Italian Civil Protection Department (Dipartimento di Protezione Civile, DPC) and thereby triggered the seismic emergency protocol. Subsequently, they elaborated the data in order to produce the first manually reviewed hypocenter, which was published on the Institute's website at 01:53:18 UTC. The sequence following this mainshock generated thousands of earthquakes in the epicentral area, which the INGV automated localization system processed and detected along with the usual seismic activity in the rest of the Italian territory. In this paper we analyze the behavior of the automated system and of the data lifecycle management procedures in such extraordinary conditions. In particular we want to measure the capability of the system to manage the huge data flow, in terms of frequency and size of seismic events and its ability to remain fairly responsive and accurate in accomplishing its duty in the expected time. This will help us to identify potential problems and to suggest necessary improvements to better serve the INGV mission for Civil Protection.*

## I. INTRODUCTION

The information system AIDA was built to collect, process, archive and distribute seismic data in near real-time. It became fully operational in May 2012, when it substituted the former main earthquake detection system at INGV. Its core components are the Earthworm software for the real-time earthquakes detection [Johnson et al., 1995], Seis-Comp3 package [SC3] for the exchange and archiving of seismic waveforms and a MySQL Database to store earthquakes data. In order to meet the specific requirements of the Institute's mission, the system features many custom modules, tools and applications devel-

oped in house. For a detailed description of the overall system architecture and a previous evaluation of its performance, refer to [Mazza et al., 2012]. Since its initial deployment, the AIDA system has been continuously developed further and gradually improved, to make it more accurate and performing. Considerable work was made to refine the software procedures and increase hardware performances, enabling the system to respond in a few seconds when triggered by an earthquake. At the same time, the load on the system has progressively increased. This is due both to the volume of data to be processed and to the constantly increasing amount of requests and queries for various types of seismic data. The complexity of the whole system justifies the ac-

tual impossibility to perform an overall test, including the human interactions with the system. Only few parts of the system, like the insertion into the database of the earthquakes data or the localization system, or the websites were tested singularly. The seismic sequence starting with the August 24, 2016 magnitude 6.0, is considered a “real life stress test” and we illustrate how this sequence and the large amount of detected seismic events impacted the whole processing and data dissemination system. We analyze various aspects in order to assess the performance of the AIDA system, and we highlight some of the strengths and weaknesses of the current system. This analysis should provide tangible actions to be proposed for future developments of the system.

## II. IMPACT ON THE PROCESSING SYSTEM

The INGV Earthworm implementation involves four different systems running in parallel to perform the event detection. Each one provides, for each earthquake detected, a package of SAC waveform files used by the software for the interactive revision [Bono, 2008]. We show in figure 1 the amount of data produced by a single Earthworm server during 2016. At the moment of this writing the September column is not complete. The monthly average data is about 74 GB during the first seven months of 2016 averaging to a rate of approximately 360 MB/day.

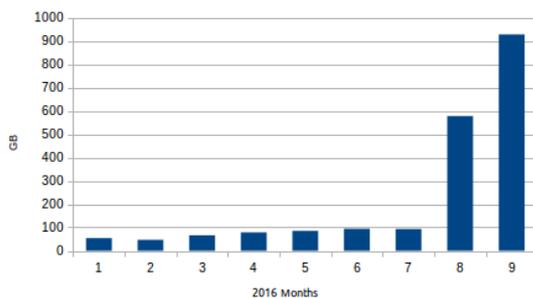


Figure 1: Data produced by a single Earthworm server.

In figure 2, we show that during the first days of the seismic sequence, the daily data rate ran up to 73GB, almost the same value of the average monthly data rate seen before. The backup routines were quickly modified to avoid risk of disk full on the various systems.

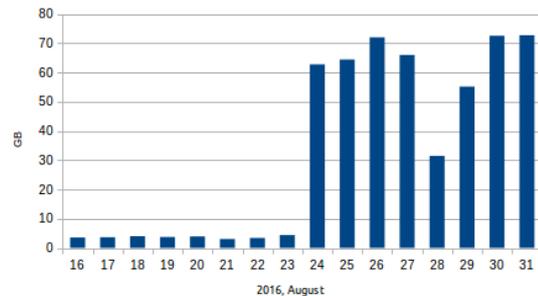


Figure 2: Earthworm data archived in August 2016.

The data volume growth needs to be seriously taken into account for future system upgrades. Our database is filled by all the seismic events data calculated by the Earthworm systems and by the manual reviewers. During the first month of the seismic sequence starting from day 24 August, we stored 220K localizations with corresponding 3.3M phases and 2.7M amplitudes. To better understand the difference with normal activity, notice that this amount of data is comparable with what was recorded by the system during the whole past year.

## III. REAL-TIME PERFORMANCE EVALUATION

The real-time processing and localization system has been properly working during the sequence. In particular, we have been able to give prompt information to our counterparts and comply with all agreements and obligations, taking also into account that in agreement with DPC the magnitude threshold for immediate phone communications was increased soon after the mainshock. To meet those obligations, the automatic results must be available before the time limits reported in Table 1.

**Table 1:** Automatic solutions communication rule

	Minutes after origin time	
	2	5
$2.5 < M < 3.0$	None	Final automatic location and magnitude
$M \geq 3.0$ (*)	Location and magnitude estimate	Final automatic location and magnitude

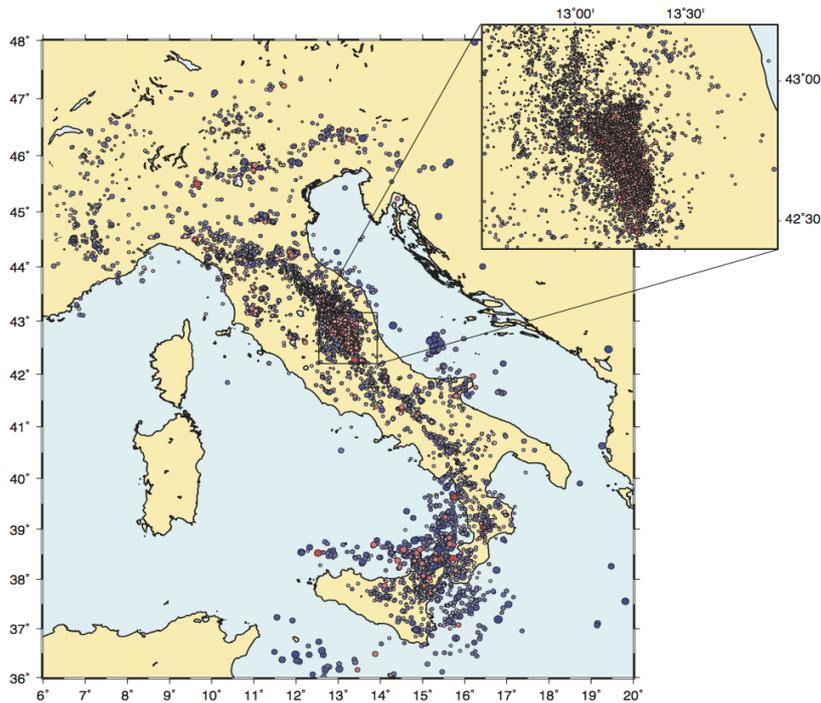
(\*) The threshold was raised to 4.0 one hour after the mainshock to limit the overload for the seismologists on duty (and the counterparts at DPC) caused by continuous phone calls.

The comparison between the number of events meeting the criteria and the number of actually recorded ones is shown in Table 2.

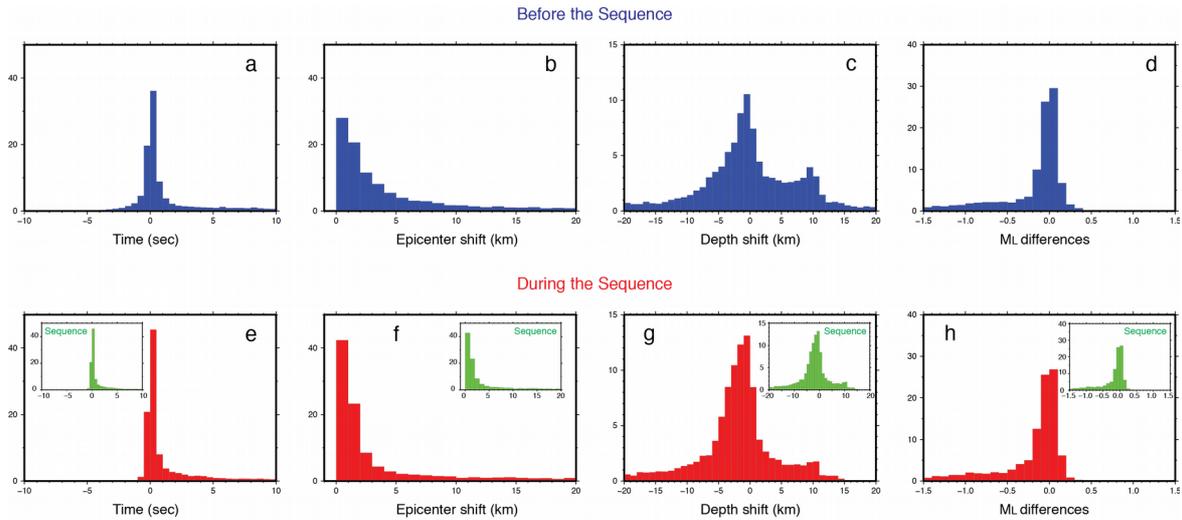
**Table 2:** Locations matching the rules

	Minutes after origin time	
	2	5
First hour	27/28(*)	28/28
After	48/49(**)	49/49

In the first hour after the mainshock, we recorded 28 quakes to be notified to DPC; only one (\*) of those locations, belonging to the sequence, was affected by an 11 seconds delay. Searching until September 22, 2016 we discovered that only another (\*\*) record, always belonging to the sequence and occurred the first day, was delayed of about 44 seconds.



**Figure 3:** Map showing the distribution of the earthquakes for which the automatic and revised localizations are compared. Red circles indicate the earthquakes (approximately 11,000) occurred in the “during-the-sequence” period, i.e. from 24th August onward. The approximately 11,000 blue circles indicate earthquakes occurred on Italian territory before that date, backward to April 2015. A zoomed view of the seismicity in the area affected by the Amatrice seismic sequence is in the top right inset of the map. Circles are scaled based on magnitude of the earthquakes.



**Figure 4:** Statistical distributions of the differences between pairs of hypocentral parameters belonging to automatic and revised earthquake localizations. In panels *a–d*, on the top row, the differences between revised and automatic (*a*) origin time, (*b*) epicenter, (*c*) depth and (*d*) local magnitude *ML* are shown, respectively, for earthquakes occurred before the seismic sequence onset (blue histograms). In panels *e–h*, on the bottom row, the differences between revised and automatic (*e*) origin time, (*f*) epicenter, (*g*) depth and (*h*) local magnitude *ML* are shown, respectively, for earthquakes occurred during the sequence (red histograms). Also, in panels *e–h*, the corresponding differences computed only for the events belonging to the seismic sequence (red circles in the top right inset of Figure 3) are shown as smaller insets for comparison (green histograms).

#### IV. ACCURACY OF THE AUTOMATIC LOCALIZATION SYSTEM

In this section we assess the quality and accuracy of about 11,000 localizations and magnitude estimates generated by the automatic system during the Amatrice seismic sequence by comparing them to the ones which were successively revised by the operator on duty (red circles in Figure 3). Also, to gain insights on the behavior of the system in such a “stress condition” we perform the same comparison on a set of as many pairs of localizations and magnitude estimates (automatic and revised) for earthquakes occurred before the onset of the sequence (blue circles in Figure 3). Even though both the sets of hypocentral parameters pairs are distributed over the whole Italian territory, those in the “during-the-sequence” period are prevalently related to the ongoing seismic sequence (see top right inset in Figure 3). In Figure 4 we compare four main hypocentral parameters, the origin time, the epicenter on surface, the depth and the local magnitude

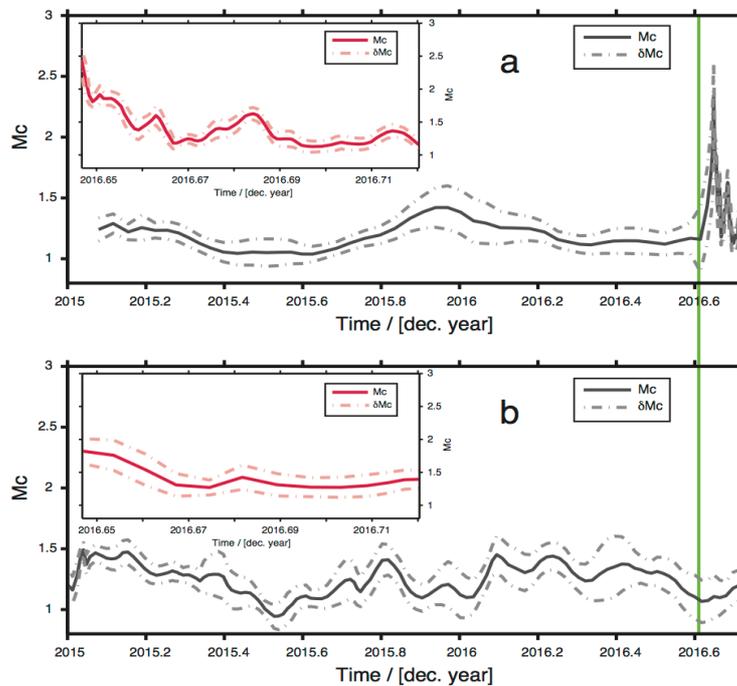
*ML* for the “before-the-sequence” and the “during-the-sequence” localization pairs. Results of this exercise demonstrate that the behavior of the system is not affected by the data load increase during the sequence. Rather, the “during-the-sequence” localizations pairs show smaller differences between automatic and revised hypocentral parameters, hence are characterized by an overall better automatic localization. Only the automatic magnitude estimate is slightly worse in the “during-the-sequence” period. This is due to the frequent presence of multiple earthquakes in the same time window, which may lead to wrong automatic associations of the maximum amplitude of the seismic signal to the right event.

#### V. SYSTEM DETECTION CAPABILITY

A critical issue to face in the aftermath of a major earthquake is the magnitude completeness of the aftershock catalog. This issue arises from the under-reporting of short-term aftershocks, especially smaller ones in earthquake catalogs,

simply because systems are not able to distinguish them in time windows containing larger events [Enescu et al., 2007]. However this “under-reporting effect” may affect the whole seismic catalog, because of possible deficiencies in recording capabilities by the system under heavy load condition. To evaluate how the heavy load of data generated by the Amatrice seismic sequence affected the detection capability of the system we compute the magnitude of completeness ( $M_c$ ) as a function of time for two different earthquake catalogs: earthquakes occurred inside the seismic sequence area (the area represented by the top right inset in the map of Figure 3); earthquakes occurred outside

this area. We perform the calculation for the two catalogs in two time windows: (1) from January 2015 to the end of September 2016; (2) from the onset of the seismic sequence (24 August 2016) to the end of September 2016. Results of these calculations are shown in Figure 5a e 5b for the “inside” and the “outside” catalogs, respectively. We compute the  $M_c$  vs time relationships for the selected earthquakes catalog on running windows of 500 events with 50% overlap (Figure 5a,b). On each sample, we determine  $M_c$  as the magnitude at which 90% of the data can be modeled by a power law fit [Wiemer and Wyss, 2000].



**Figure 5:** (a)  $M_c$  as a function of time for the “inside-catalog” from January 2015. Continuous dark gray line represent the  $M_c$  values computed on running windows of 500 events, dashed gray lines indicate the standard deviation. In the top left inset of panel (a), the continuous red line represents the  $M_c$  vs time for the “inside-catalog” when only data from the start of the sequence (24 August 2016) are considered in the calculation, dashed red lines indicate its standard deviation. (b)  $M_c$  as a function of time for the “outside-catalog” from January 2015. In the top left inset of panel (b) the  $M_c$  vs time for the “outside-catalog” is shown, when only data from the start of the sequence (24 August 2016) are considered in the calculation. Symbols and color in panel (b) have the same meaning as in panel (a). The green bar marks the time of occurrence of the ML 6.0 Amatrice earthquake. Calculations shown in this Figure are made using the ZMAP code [Wiemer, 2001].

Uncertainties on the  $M_c$  values were calculated by bootstrapping each sample with 200 realizations, and indicated with dashed lines in Figure 5a,b. We observe that, before the  $M_L$  6.0 Amatrice earthquake in the sequence area (inside catalog, Figure 5a), the  $M_c$  is always below 1.5, with upward oscillation due to the higher weather-related noise level in the winter months. The  $M_c$  rises to 2.7 immediately after the main shocks (green bar in Figure 5), and then decreases again below the 1.5 threshold in few days. This can be observed in better detail looking at the red line in the inset of Figure 5a, where only data from the start of the sequence (24 August 2016) are considered in the calculation. The trend of the  $M_c$  for the “outside-catalog” is characterized by more oscillations, related to the existence of multiple different conditions in an area as large as the whole Italian territory (proximity to the coasts, anthropic-related noise, weather-related noise, etc.), even though it remains always below the 1.5 threshold. No significant variations are observed at the time of the occurrence of the  $M_L$  6.0 Amatrice earthquake (green bar in Figure 5) when data from January 2015 are considered. A slight increase of  $M_c$  up to 1.8 is observed for few days if only the data from the start of the sequence (24 August 2016) are considered in the calculation (red line in the inset of Figure 5b). Summing up, we can conclude that the heavy load of data generated by the Amatrice seismic sequence did not significantly affect the detection capability of the system, neither inside the area affected by the sequence nor in the whole Italian territory.

## VI. DATA SHARING AND DISSEMINATION

The CNT website (<http://cnt.rm.ingv.it>) is our main seismic parametric data sharing and earthquake information portal. It received more than one million contacts the day of the mainshock. See Figure 6 for a more detailed time series. Our hosting provider blocked the traffic immediately after the sudden increase of connections only few minutes after the main-

shock, assuming that this amount and pattern of http requests was corresponding to a distributed cyber attack.

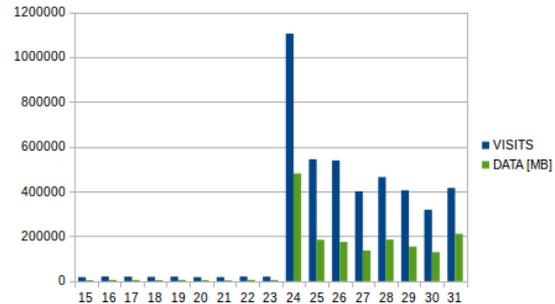


Figure 6: Connections to CNT web site.

The web portal *ISIDe* (<http://iside.rm.ingv.it>) [Mele et Al. 2016] is another instrument for data dissemination targeting more specifically users from the research community. Although it registered more than 50,000 accesses shortly after the mainshock, accessibility was not affected. We experienced very good performances, and many new users (around 25% of all contacts) were able to connect, to register to the portal and to browse and request data. See Figure 7 for more details.

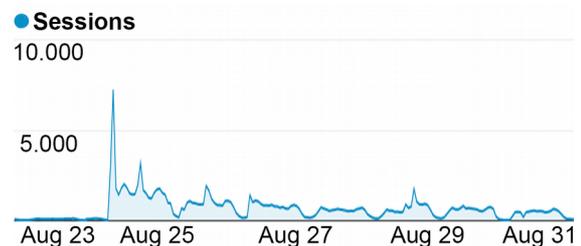


Figure 7: Contacts to ISIDe web portal.

Our webservice server (<http://webservices.rm.ingv.it>) provides methods to access raw data and data products programmatically through standardized and public APIs. These data include parametric event data, station information and metadata, as well as raw seismogram waveform data. Parametric earthquake data like events, magnitude, phases and picks are distributed in the standardized QuakeML format (<https://quake.ethz.ch/quakeml/>), a flex-

ible, extensible and modular XML representation of seismological data.

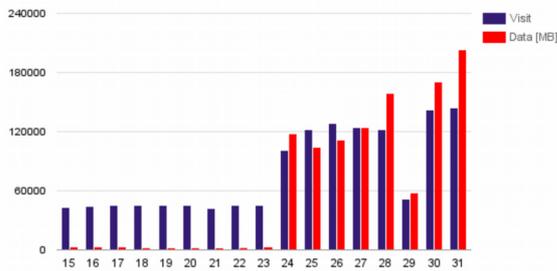


Figure 8: Data downloaded via web-services.

During the Amatrice earthquake, the number of requests and the amount of downloaded data are increased dramatically. Nonetheless, the service remained always available and guaranteed access to the requested data. This service is also used to provide the input data in JSON format (<http://www.json.org>) for INGV's mobile applications (Apps) called *INGVterremoti* for the iOS and Android operating system. The number of requests made by these Apps increased from a value of about 2,000 from the previous day to almost 150,000 on August 24, 2016. The development of the traffic can be observed in Figure 8. All these figures (Fig. 6 - Fig. 8) show an immediate increase of sessions and traffic in the night of August 24, 2016, followed by a slow and constant decrease over the following days.

## VII. CONCLUSIONS

The 2016 Amatrice earthquake sequence has severely tested our automatic and interactive processing systems. It generated an heavy load of new data in a relatively short time period, a load which was several times larger than usual in terms of number of events, archived data, bandwidth and requests by users and automated processes. Nevertheless, the system's behaviour and response was satisfactory, in terms of event processing speed, detection capability and accuracy, and service uptime and responsiveness, although the system needed some extra work to remain efficient without

running out of storage space. This experience teaches us that we need to continuously upgrade the hardware and notably disk space, in order to keep up with the constant growth of the seismic networks and constantly improve detection capabilities. Moreover we should try to reduce the amount of data written by the system, reducing or completely eliminating the use of SAC waveforms during the manual revision in favour of the use of time series web-services like the IRISWS-timeseries service (<http://service.iris.edu/irisws/timeseries/1>). The procedures for the insertion of seismic data into our database systems, even if satisfactory, would benefit from further improvements, in particular some fine tuning of the database server and data insert optimizations should be done, in order to obtain even better performances. Quality of automatic magnitudes during the sequence is slightly worse than in the usual scenario. This will be further investigated later on, as a finer tuning of the time window used to search the maximum amplitude may arguably guarantee some improvement in the automatic calculation of the magnitude, limiting the cases of wrong associations of amplitudes to seismic signals. Dissemination of information and data to the public has been very successful with millions of request fulfilled by our websites. However, new and improved solutions for even more requests and higher data volumes should be prepared and established, because we have to anticipate the continuous growth of the Internet population over the coming years.

## Acknowledgements

This study benefited from funding provided by the Italian Presidenza del Consiglio dei Ministri, Dipartimento della Protezione Civile (DPC); scientific papers funded by DPC do not represent its official opinion and policies.

## REFERENCES

[Bono, A. 2008] *SisPick! 2.0 Sistema interattivo per l'interpretazione di segnali sismici*, Rapporti tecnici INGV, n.59

- [Johnson et Al., 1995] Johnson, C. E., A. Bittenbinder, B. Bogaert, L. Dietz, and W. Kohler (1995). *Earthworm: A flexible approach to seismic network processing*, Incorporated Research Institutions for Seismology (IRIS) Newsletter 14, 1–4
- [Enescu, B., J. Mori, and M. Miyazawa, 2007], Quantifying early after-shock activity of the 2004 mid-Niigata Prefecture earthquake (Mw6.6), *J. Geophys. Res.*, 112, B04310, doi:10.1029/2006JB004629.
- [Wiemer, S., and M. Wyss, 2000] *Minimum magnitude of complete reporting in earthquake catalogs: Examples from Alaska, the western United States, and Japan*, *Bull. Seismol. Soc. Am.*, 90, 859–869, doi:10.1785/0119990114.
- [Mazza et Al., 2012] Mazza S., Basili A., Bono A., Lauciani V., Mandiello A., Marcocci C., Mele F., Pintore S., Quintiliani M., Scognamiglio L. and Selvaggi G. (2009). *AIDA – Seismic data acquisition, processing, storage and distribution at the National Earthquake Center*, INGV, *Annals of Geophysics* (2012), Vol. 55, No. 4.
- [Mele et Al., 2016] ISIDe working group (2016) version 1.0, DOI: 10.13127/ISIDe
- [QuakeML] Danijel Schorlemmer, Adrian Wyss, Silvio Maraini, Stefan Wiemer, and Manfred Baer: *QuakeML - An XML schema for seismology*, ORFEUS Newsletter, volume 6, no 2, October 2004 (HTML)
- [SC3] <http://www.seiscomp3.org>
- [Wiemer, S., 2001] *A software package to analyze seismicity: ZMAP*. *Seismological Research Letters*, 72(3), 373-382.