Istituto Nazionale di Geofisica e Vulcanologia

INGV Seismic Waveform Distribution and Archive Integration.

Lauciani V., Mandiello A.G., Marcocci C., Pintore S., Mazza S.

Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata 605, 00143 Roma

Abstract

The present, continuously increasing, number of seismic stations is posing a serious challenge to data centers regarding the management of data flows and station information. CNT is following the approach of a distributed data archiving, participating to the European Integrated Data Archive (EIDA), as proposed and realized in the European Project NERIES.
Data are collected and archived in a Storage Area Network with a SEED Data Structure, while station information is organized in a MySQL database. Using the ArcLink protocol (by the Geofon Group, GFZ) station information is integrated in one virtual repository. Similarly, seismic data are shared in a common data-bank. This approach results in a highly scalable archive, in which a strong backbone is provided by the major Institutions, but where also minor participants can play their role. Each data producer can manage data and station information from its privileged workplace, whereas the technicalities behind the scene are hidden to the data users, who can access the same integrated data set contacting one data center only.
As partners of EIDA we have given access to the European Archive through our web pages at http://eida.rm.ingv.it, where data from a huge number of global, regional, national and local broad band stations are available. At the same time, non-restricted data provided by projects and cooperations are available in the same page, as well as data from national and international temporary deployments. Automatic data quality control is at an experimental stage. Tools for non-interactive data extractions are under development.

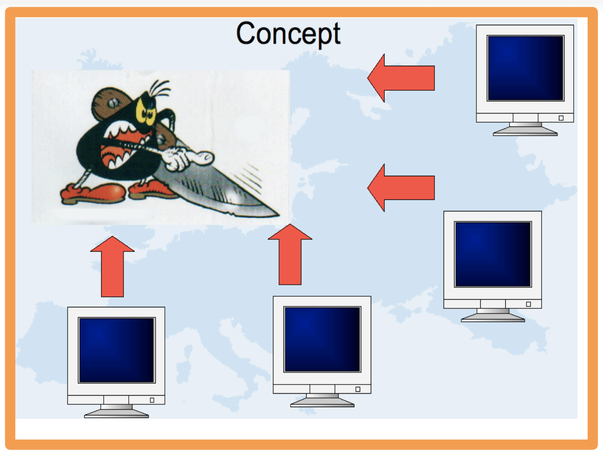
Archiving Broad band data

"Broad-band seismic recording has the purpose of recording all seismic signals of interest in a single three-component set of seismograms, as opposed to recording different types of signals in separate channels. " (E. Wielandt)

This implies:

- Continuous signals
- Accurate metadata
- Multiple data streams (desampling)

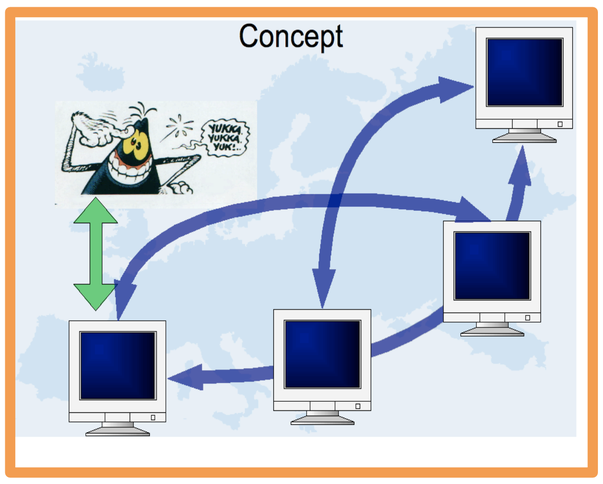
(but event data are still welcome)



Data distribution

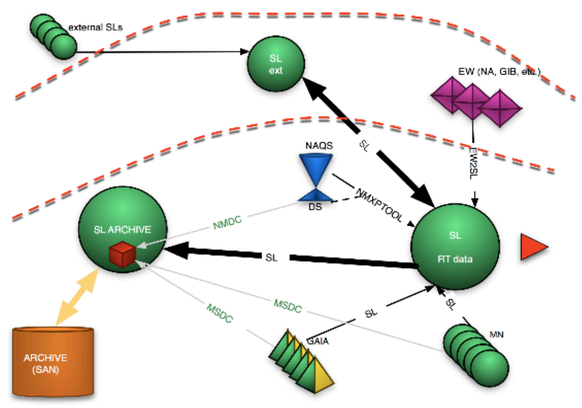
When the station number is relatively small and the data producers not many, the viable choice is one Data Center. Mail/ftp instruments like Spyder, BreqFast, AutoDRM are satisfactory.

As stations and data increase keeping track of them is a challenging task, for both data centers and users, because data are located in many places and station information is difficult to mantain. A distributed approach is more suitable. NetDC is the first attempt, still mail/ftp based.



INGV-CNT Data Archive

- Data are collected from different sourcesby a SeedLink server.
- Formatted
- Desampled
- Quality checked
- Archived in a Storage Area Network with a SEED Data Structure.
- Station information is organized in a MySQL database.



ArcLink protocol

The generic request format is following:

REQUEST <request_type> <optional_attributes>

<start_time> <end_time> <net> <station> <stream> <loc_id> <optional_constraints>

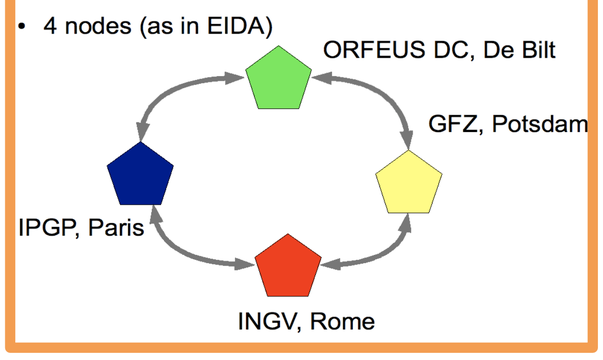
[more request lines...]

END

3 types of REQUESTs: inventory, waveforms, responses

Distributed Archive Configurations

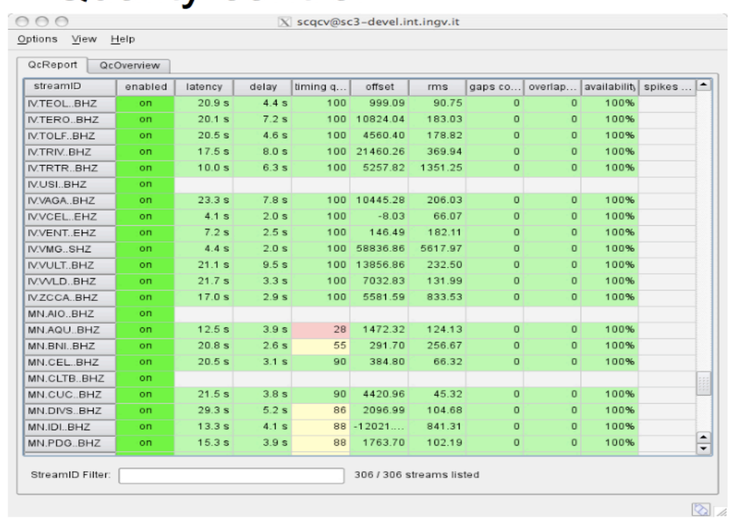
- 4 nodes (as in EIDA)



Quality control

SCQC monitors in real time:

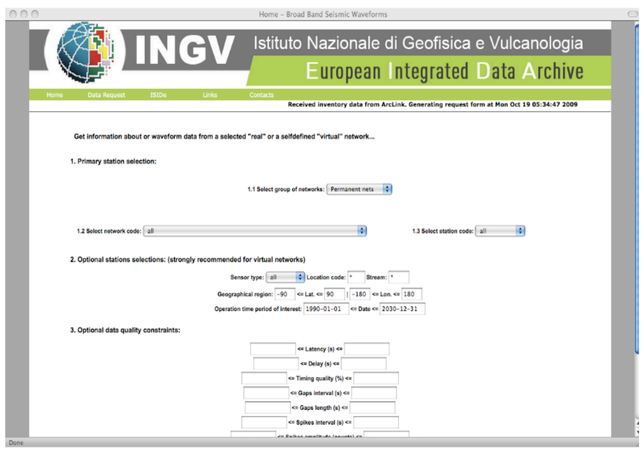
- Latency
- Delay
- Time quality
- Offset
- Rms
- Gaps
- Overlap
- Availability
- Spikes



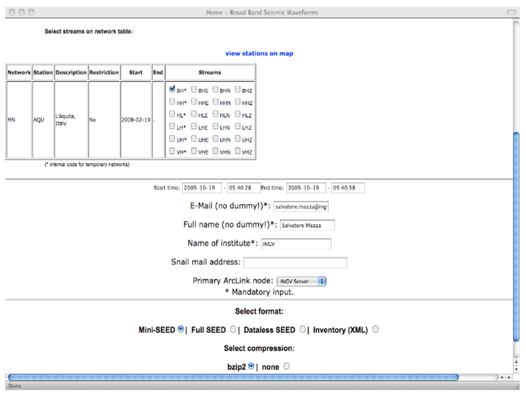
streamID	enabled	latency	delay	timing q...	offset	rms	gaps co...	overlap...	availabilit	spikes ...
IV.TEOL..BHZ	on	20.9 s	4.4 s	100	999.09	90.75	0	0	100%	
IV.TERO..BHZ	on	20.1 s	7.2 s	100	10824.04	183.03	0	0	100%	
IV.TOLF..BHZ	on	20.5 s	4.6 s	100	4560.40	178.82	0	0	100%	
IV.TRIV..BHZ	on	17.5 s	8.0 s	100	21460.26	369.94	0	0	100%	
IV.TRTR..BHZ	on	10.0 s	6.3 s	100	5257.82	1351.25	0	0	100%	
IV.USI..BHZ	on									
IV.VAGA..BHZ	on	23.3 s	7.8 s	100	10445.28	206.03	0	0	100%	
IV.VCEL..EHZ	on	4.1 s	2.0 s	100	-8.03	66.07	0	0	100%	
IV.VENT..EHZ	on	7.2 s	2.5 s	100	146.49	182.11	0	0	100%	
IV.VMG..SHZ	on	4.4 s	2.0 s	100	58836.86	5617.97	0	0	100%	
IV.VULT..BHZ	on	21.1 s	9.5 s	100	13856.86	232.50	0	0	100%	
IV.VVLD..BHZ	on	21.7 s	3.3 s	100	7032.83	131.99	0	0	100%	
IV.ZCCA..BHZ	on	17.0 s	2.9 s	100	5581.59	833.53	0	0	100%	
MN.AIO..BHZ	on									
MN.AQU..BHZ	on	12.5 s	3.9 s	28	1472.32	124.13	0	0	100%	
MN.BNI..BHZ	on	20.8 s	2.6 s	55	291.70	256.67	0	0	100%	
MN.CEL..BHZ	on	20.5 s	3.1 s	90	384.80	66.32	0	0	100%	
MN.CLTB..BHZ	on									
MN.CUC..BHZ	on	21.5 s	3.8 s	90	4420.96	45.32	0	0	100%	
MN.DIVS..BHZ	on	29.3 s	5.2 s	86	2096.99	104.68	0	0	100%	
MN.IDI..BHZ	on	13.3 s	4.1 s	88	-12021....	841.31	0	0	100%	
MN.PDG..BHZ	on	15.3 s	3.9 s	88	1763.70	102.19	0	0	100%	

StreamID Filter: 306 / 306 streams listed

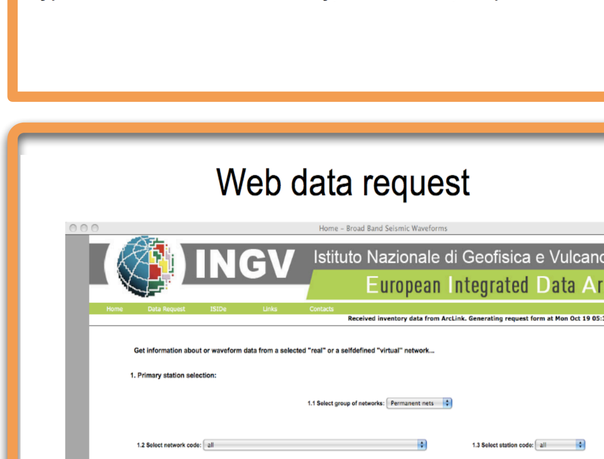
Web data request



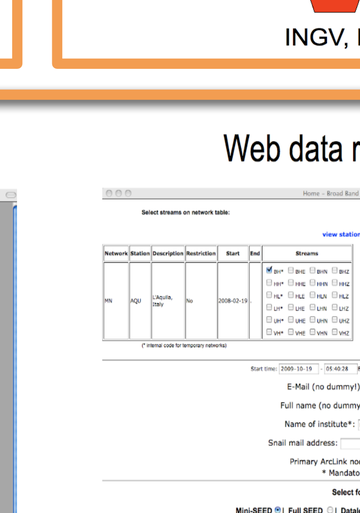
Web data request



Quality control



Power Sperctral Density time evolution

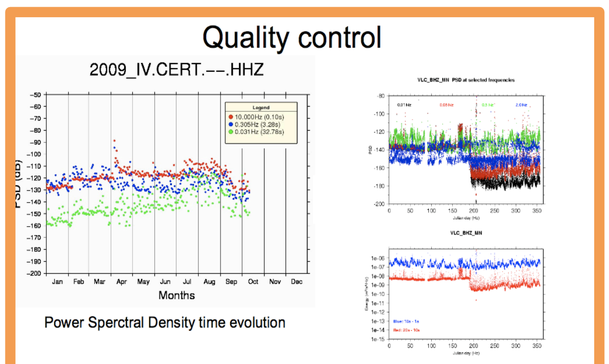


Sleeman, R., and J. Vila
Towards an automated Quality Control Manager for the Virtual European Broadband Seismograph Network (VEBSN)
Submitted, Orfeus Electronic Newsletter, 2005

Homework

- Recover MN 1990-2005 data
- Include IV data before 2008
- Tools to "batch" requests
- Web Services

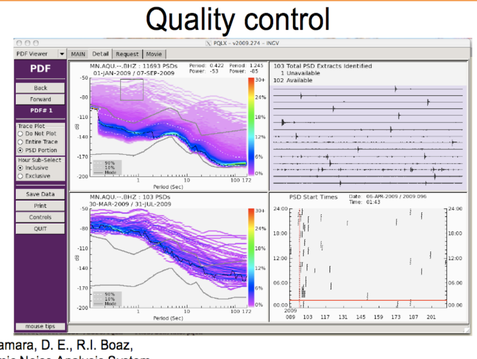
Quality control



Power Sperctral Density time evolution

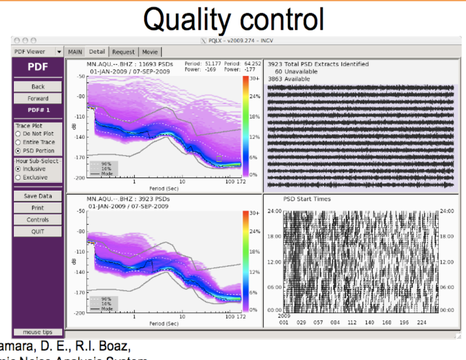
Sleeman, R., and J. Vila
Towards an automated Quality Control Manager for the Virtual European Broadband Seismograph Network (VEBSN)
Submitted, Orfeus Electronic Newsletter, 2005

Quality control



McNamara, D. E., R.I. Boaz,
Seismic Noise Analysis System,
Power Spectral Density Probability Density Function: Stand-Alone Software Package,
United States Geological Survey Open File Report, NO. 2005-1438, 30p., 2005.

Quality control



McNamara, D. E., R.I. Boaz,
Seismic Noise Analysis System,
Power Spectral Density Probability Density Function: Stand-Alone Software Package,
United States Geological Survey Open File Report, NO. 2005-1438, 30p., 2005.