From seismic stations to integrated datacenters and computational facilities



Javier Quinteros, Angelo Strollo and the GEOFON team





The GEOFON mandate

GEOFON aims at providing scientific infrastructure in the form of global seismic network, data archival and monitoring facilities to facilitate cooperation in seismological research leading to a better understanding of our complex system Earth.

It pursues these aims by operating and maintaining a global network of permanent broadband stations in cooperation with local partners, facilitating real time access to data from this network and those of many partner networks and plate boundary observatories, providing a permanent and secure archive for seismological data.





The GEOFON mandate

Using real-time data streams, GEOFON determines rapid automatic location estimates for all globally recorded earthquakes and most regional ones, and provides manually revised solutions for the largest earthquakes with minimal delay. This service provides basic rapid earthquake information to earthquake and tsunami warning centers worldwide, governmental agencies, disaster management teams, news media and scientists at the GFZ and elsewhere.





Describing an Earthquake

An earthquake's most fundamental characteristics are

Location

- *Hypocentre* or *focus*: Geographical location of an earthquake i.e. its epicentre and depth.
- *Epicentre*: Point on the Earth's surface above the hypocentre.
- Origin time
 - Start time of the rupture.
- Focal mechanisms
 - Fault geometry: Geometric orientation of an earthquake's fault, dependent upon the local geology/tectonics.
 - *First motion:* Identifiable from the seismograms, used to identify the fault geometry
- Size

GFZ

Helmholtz Centre

- *Magnitude* or *moment:* Value describing the strength of an earthquake.





The GE global seismic network and GEVN

~80 GE stations, ~30 GFZ stations, ~800 stations for GEVN







Global earthquake monitoring

Since January 2011 also MT solutions for events larger than M 5 (depending on data quality) 6085 events and 731 MT solutions published in 2012







GEOFON in numbers

•81 GE stations + 38 cooperated stations (affiliated) •~900 stations acquired and processed in real-time •~100 TB/year real-time data distributed via seedlink -3000 stations ~70 TB in archive today •4.2 million customized requests in 2014 •~6 TB/year of data shipped from archive •~6500 published events in a year •~1000 published moment tensors in a year •~28000/day unique connections to the web pages (average) 2014)

-330 SeisComP3 free non-commercial licenses

GEOFON data, products and software are open!





Early Warning Systems





Tsunami generation









Tsunami propagation







The 26th December 2004 - Banda Aceh













The Challenge, Tsunami Warning for Indonesia •Earthquake sources very close to affected coasts •Tsunami travel time 20 - 40 minutes



require tsunami warning within \approx 5 minutes





The 2014 Iquique Earthquake

From rapid event information to scientific activities through HART deployment and data curation Time

- 2014.04.01 23:46:49 UTC Event OT
- + 11 minutes 1st automatic solution M 7.9
- + 55 minutes 1st MT manual solution Mw 8.0
- + 1 day MT manual review Mw 8.1
- + 1 day Event special page (DOI)
- + 1 day HART Coordination meeting
- + 10 days Deploying HART instruments
- + 5 months Nature paper and more
- + 12 months DFG proposal submission
- + 16 months HART data collected/archived
- + 18 months DFG proposal accepted







Rapid Earthquake Information







Scientific use of data





Imaging with earthquake data

□ Moho, Lithosphere, Transition zone, Anisotropic structure □Example: Pamir-Hindukush continental subduction (4 sections, 2 departments directly involved)

Helmholtz Centre POTSDAM



a)TIPTIMON

c) Joint and relocated

OCIATION

GFZ scientific infrastructure

The GFZ seismological data archive: 70+ TB, 3000+ stations





http://geofon.gfz-potsdam.de/waveform http://eida.gfz-potsdam.de



EIDA/Orfeus Construction of a European Federation





EIDA within Orfeus







EIDA within Orfeus

Protocols to share data
Arclink protocol
Synchronization of metadata

•Daily

Clients and tools development

SeisComP3

•Common policies for f.i. data curation

•Statistics

Maintainance of Routing Tables





Seismic Waveform Data

How to get waveforms and metadata: WebDC3 portal



http://eida.gfz-potsdam.de





EIDA/Orfeus New Projects: EUDAT, EPOS





Seismic Waveform Data EUDAT2020 The pan-European Data Infrastructure

EUDAT offers common data services, supporting multiple research communities as well as individuals, through a geographically distributed, resilient network of 35 European organizations.

GEOFON involvement:
Safe replication
Identity Management
Data Discovery
Dynamic Data





http://eudat.eu/





Seismic Waveform Data EPOS-IP European Plate Observing System

EPOS is integrating the diverse, but advanced European Research Infrastructures for solid Earth Science, and will build on new e-science opportunities to monitor and understand the dynamic and complex solid-Earth System.







GFZ scientific infrastructure

Enhancing accessibility at European and Global scale

□GFZ seismological datasets are open data (except embargo for temporary experiments)

EIDA (European Integrated Data Archive)
=> EUDAT (The pan-European Data Infrastructure)
=> EPOS (European Plate Observing System)





Safe replication Identity Management Data Discovery Data Staging Dynamic Data Integration and interoperability with other solid earth science infrastructure





EIDA Next Generation Software

More than just data and federated archive

Coordination of data holdings and software/strategic developments
Provides quality control of data/metadata
Helps define seismological center best practice' for ORFEUS community

	EDA		Access	to EID	A Da	ta Are	chives		Orfe	us
										Help
Explore events	Explore st	ations	Sabert	request		() ow of	ioad data		View console	
NOR'S CONTROLS.	12	ENERTAND D	TACION HAP							ø
Ser Information Browns Investory - Un descript an from <u>1980</u> to <u>2010</u> enack Types 11 nots enack Code. 11 Detective	er Saplet	de can	and the second sec	THE REAL				2 0	HJRAZA Lana Paganan Higheren	Sec
 temperary network; + tem by Code: 3e Region; Filer stations by station 	te Grents	Vier des Bill		Algenta	- and			and the second	1.	
- bergensky hatersky o cene betens be Code De Region Filme ataliers by atalier All Stationer	Ar Events	Providence Color Lines heft Dirig Director Andre Color	TANDE LIT	Agent	a agent			and and	7 .	
 bergensky haltersky s cene be Code De Region Filme statiers by statier All Statiene 	in Conta Internet	Provide Rel Unit and Del Decent Ann II Record C	TAKON URT	Algera Rigera	a ragana	Delute	Datare	Save Stat	Laguer	
 temporary networks a centre antenne. be Code: the Region. All: Statistics by disker All: Statistics temporary be Code: the Limit Code: the Co	ta Conta In Conta In I	Prevant Anno 197 Des des des Anno 197 Des des des Anno 197 Des des des des des des des des des des d	riteration FT + drag max RCONULT 74 events)	Algerta No to color	Annual Annual	Delute	Durane I	Save Stat	Lagrand	
 Interpretation and enables of energy and enables of the end of the second second second	in Gootty andre andre allen al descretu screditte	The data and the last	74 events) n Taxe	Napata an Ta anka Napata Ta anka	Frances Frances Frances	Deluta 23-18	Dators Long 33.32 47.31	Save Stat Depth 8.3	Legend Tree Dable Lee Region Rethers and Gentral Lee Dables	
• temporary networks - con- constructions befores by Code the Region Filter stations by station All Staticons freezes by Code the Size Consected desired set (set 1967 and C201.16) set	A Good of Control of C	The des EX Use left Disp Excel All C Excel All C Excel All C Congo 2 2015-1 2 2015-1	74 events) n Terre 04-20101 (1-2-2) 05-20101 (1-2-2) 05-20101 (1-2-2) 05-20101 (1-2-2) 05-20101 (1-2-2) 05-20101 (1-2-2)	Agents on 12 october Mage 10 5.4 10 5.4 10 5.4 10 5.4 10 5.4	Fragers Fragers Fragers rs rs rs	Delate 33-18 36-28 36-73	Etatione Statione S3.32 -47.31 -46.29	Save Stat Days M3 M3 M3 M3 M3 M3 M3 M3 M3 M3 M3 M3 M3	Lagend Lagend The Dable Lee Marten and Dathore Diathore	
Interprete setures: encode Interprete setures Interpret	national an Events and an antisy of decodes relevants	Personal and the second	74 events) n Texes 04 2010 1011 05 2010 101 06 2010 101 00 06 2010 101 00 06 2010 101 00	Agents on 12 stalls Hegs 10 3.4 15 3.4 15 3.2 10 3.2	France France France France France	Delate 23-18 36-28 36.71 19.10	Etations Etations 53-32 -47-31 -46-29 -44-31	Save Stat Days 5.3 5.3 5.3 5.5 2.3 2.3	Legend Legend Corre Dable Lee Marthers and Correct lear Oldahome Oldahome Ungan blands	
Interprete setements Interprete Interp	national in Events and a in in initial initia initi initia initia initia initia initia initia initia initia	They don they take left Brid EVENTABLE Toronts (7 2013- 2013- 2013- 2013- 2013- 2013- 2013- 2013-	74 events) 64 events)	Napata an 14 and 40 Napata 18 and 40 Napata 18 and 40 18	Angers An	Databas Databas 33.18 36.28 36.71 19.10 8.29	Eattore Eattore 53.32 -47.31 -42.39 -44.39 -44.39 -44.39	Save Rat Days 10 10 10 10 10 10 10 10 10	Annee Annee	
Interprete statute - e- ence Interprete Interpret Interprete Interprete Interprete Interprete	Age and de	New data that the left left Respectit Press (27 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015- 17 2015-	7 4 conto 1000 UUT 7 4 conto 1000 UUT 7 4 conto 100 cont	Napata an 12 an Ma 12	Frances Franco	Delata 25-18 36-28 36-71 29-10 8-29 -37-07	Rattors Rattors 1000 1000 1000 1000 1000 1000 1000 10	5ava (24) 5ava (24) 54 54 54 54 54 54 54 54 54 54 54 54 54	Legend Legend Determine Calification Calific	
Interpret, solution, e encode In Code Se Region. All Status Ne data All Status Ne data All Status Ne data Volume Net Second Children No Code Second Children No. Reset	Append	Marchan Ray Lane Ray Dire Marchan Ray Lane Ray Dire Marchan Ray Marchan Ray Ma	74 events) 74 events) 8 Terr 90 - 2010 1 01 0 90 - 2010 1 0 90 - 2	Napara an 12 anko Maga 12 anko 12 anko 13 anko 14 anko 15 anko 15 anko 15 anko 16 anko 17 anko 18 anko 18 anko 18 anko 18 anko 18 anko 19 anko 10 anko	Franka Fr	Delata 33-18 36.71 29.38 8.29 -37.67	Eurore Eurore 53.52 -47.51 -42.53 -42.43 -42.43 -12.44	5ave Bat 5ave Bat 53 53 53 53 53 53 53 53 53 53 55 55 55	Lagord Lagord The Lagord Lagord Martheon and Galak-tome Onlati-tome Virgin Islende Perfana-Coniel National Perfana-Coniel Ros Barros Coniel Co	
Interpreters softwark, etc. Interpreters Intereadmeters Interpreters Interpreters Interpreters I	Age and	New data for una lati bio Personali	74 County 74 County 74 County 74 County 75 County 75 County 76 County 76 County 76 County 77 County 77 County 78 Cou	Napana an 1a an ka Plang 10 3.4 10 3.4 10 3.2 12 3.3 13 4.3 14 4.3 15 4.3 15 4.3	Frances Franco	Dolute Dolute 33-18 36-28 36-71 29-18 8-29 -37-67 Long	East one 53.32 -97.31 -98.29 -97.34 -72.34 0/#.59	Save Rat Depti 43 33 33 33 33 33 33 33 33 33 33 33 33	Legend Legend Control Legend Control	
Interport solution, e encode Interport I	surdar in Conta in Conta in Conta in in in in in in in	The set of	PT - ang nou PT - ang nou NCOULT 74 events) n Tore 0- 2770 - 84 0 0- 2770 - 85 0 0- 2770 - 75 0	Napara an 1a an ka 1a	Freeze Transcore Transcore Transcore Transcore Transcore Transcore Transcore Transcore Transcore Transcore Transcore	Delute Delute 33.18 36.71 19.18 8.29 -37.67 Long	Etatione 53.32 -47.31 -48.29 -42.43 -42.43 -75.34 0/# Str	524 19 19 19 19 19 19 19 19 19 19 19 19 19	Legen Construction Construct	
Interpreters softwark: - e- encode So Code So Region Anal Statutions by addee So Code So Region More Code Software	Appendix	Tree de la Constantia d	74 events) 74 events) 100 201011 74 events) 100 20101 201 20101 201 20101 2010 2010 20101 2010 20101 2010 20101 2010 2010 2010	Name an 12 and 12 and 1	Frances Franco	Delate 23-18 36.71 29.18 36.71 29.18 8.28 -37.67 teng	Europe Europe 53731 -0239 -0239 -0239 -0239 -07334 0/# 507 0	Care Dat Days 10 10 10 10 10 10 10 10 10 10 10 10 10	Annual Control of Cont	
Interport solution, or encode Interport for Code for Region All Statution Forder Interport forder	Appendix	Research for Loss left for Research for Constant for Research (Constant) Research (2015-4 Research (74 events) 74 events) 75 events) 75 events) 75 events) 76 eve	Hand A Register and Tal Starks Tal Starks Ta	France France France Transe Tr	Delute Delute 33.13 36.21 19.16 8.25 -37.07 Long 30.27 30.27	Etatione Etatione Long 42.3 42.3 42.4 4.2.4 4.2.4 1.7 2.34 0/# Str 0 0	5 ave (ha) 5 ave (ha) 5 a 5 a 5 a 5 a 5 a 5 a 5 a 5 a 5 a 5 a	Autom Au	





EIDA NG: Why?

Challenges:

Growth: archive volume, data variability, new nodes, more users, downloads

More complex user requirements

Requirements:

- Scalability: beyond traditional seismology
- Interoperability / Compliance with standards
- Extensibility: services can be migrated to other communities
- Comprehensive handling of restricted datasets
- Extended search requests
- Combining QC / SoH information with data requests
- Data replication and identification: Serving user only the best datasets





The EIDA Routing Service

- What does a routing service? Provides routing to data (streams) Routing to services Routing priorities Additional parameters information being discussed: Geolocation of data and services
- Type of archive (master, validated/non validated copy, etc)

Contact information



Can also be deployed as a standalone router to run on the client side to create virtual DCs

First stable release 10.2014





The EIDA Mediator (design phase)

The Mediator will allow users to perform complex/filtered data requests based on a number of parameters available from a number of web services. Initially developed as a server/node side application, afterwards also as a client application.







The EIDA AAI System (prototype)

Separate authentication from data services (leaving just authorization to data services).

Shibboleth

oAuth

etc.

 Support multiple authentication mechanisms (Shibboleth oAuth, E-Mail, etc.).

Secure authentication in scripts and browser.
Pattern-based authorization (data access rules).

FIDA

AUTH WS





FDSN WS

FDSN WS

FDSN WS



"OLD" Challenges from the seismological community

 Standard Waveform Data Format merging of different data sets
 Standard (Real-Time) Data Exchange Protocol Virtualization of Seismological Networks
 Standard Data Archive Access Protocol Virtualization of Seismological Data Archives
 Synchronization of distributed data sets
 Combining data and HPC services





Challenges from the seismological community

To be solved within EUDAT2020 and EPOS projects? (Hopefully!)

Dynamic Data

- \checkmark Waveforms most probably will change only during the first days.
- ✓ Seismology seems to be in a very good position compared to other sciences

• PID minting for waveforms

 \checkmark Time resolution? At the day level for every stream.

Reproducibility of experiments

- ✓ Freezing datasets in time
- ✓ Millions of requests referencing thousands of records
- ✓ Before and after curation of data/metadata
- ✓ And real-time data?





Challenges from the seismological community

To be solved within EUDAT2020 and EPOS projects? (Hopefully!)

- Sharing datasets with the world
 - ✓ Same problems as the reproducibility
 - ✓ What about real-time data?
- Integration with other Earth sciences
 - ✓ Not enough standards in other branches of science
 - ✓ Different levels of evolution regarding infrastructure
 - ✓ Protocols
 - ✓ Data formats
 - ✓ Semantics! What does f.i. "deptht" means?





Thanks for your attention!



