LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Seismological Software Developments at LMU Munich

ObsPy & SeisHub

Robert Barsch & Tobias Megies

Table of Contents

Table of Contents

SeisHub

Native, document-centric XML database

- RESTful Web service (HTTP, HTTPS),
- Standard relational database as back-end
- Both worlds: SQL for querying and manipulating data and any standard connected to XML, e.g. XSLT or XSD
- Not restricted to seismology at all

Extended to a “classical” seismic database

- Index of local, file-based waveform archive (MiniSEED, GSE2, SAC, ...)
  - Meta Data: Gaps, overlaps, quality and timing information
  - Waveform previews (30s)
- XML resource types for handling inventory data (XML-SEED) and events (QuakeML based)
- Remote waveforms access (ArcLink)

SeisHub: Technical Details

- Python-based, standalone web service
- Platform independent, open source (GPL)
- Implementation of various web protocols, like HTTP, SSH, SFTP
- Plug-in architecture: Dynamic discovering and loading of modules and support for Python .egg files
- Development remarks:
  - Test-driven development proven software, so far → ca. 250 test cases
  - Well-documented source code
  - Subversion
  - Trac: ticket system and project wiki

SeisHub: Database Design

Data storage

- Primary data → file system
  - Continuous waveform archive (MiniSEED, GSE2, SAC ...)
  - Other data via (GeoTIFF, GPS time series, etc.) file system
- Meta Data → Web service on top of a XML/relational database hybrid
  - Data is packed into a XML document → Data structure is within the document, no need for a predefined database schema
  - XML resources are archived into a BLOB field
  - Only searchable values are indexed
  - Pointers to primary data

SeisHub: Database Design

Data access

- HTTP/HTTPS: REST web service
  - XML documents have a fixed resource identifier (URL's)
  - Data transformation via XML Style Sheets on request (?output=...)
  - Data validation via Schema (XML Schema, RelaxNG, Schematron) on resource upload
  - Document properties like related meta data or indexes
- SFTP: XML documents mapped into a virtual file system

SeisHub: Database Design

- Indexing
  - Generated using a XPath expression, type and additional options
  - Simple creation + reindexing via web interface
  - Various build-in types (datetime, bool, numeric, double, float, etc..)
  - ProcessorIndex: custom processing
- Searching
  - XPath-like query on XML catalog object (restricted to indexes)
  - SQL on database object
- Mapper: predefined queries & output format bound to an fixed URL
- FileSystemResource: integrates a file system directory (read only)

SeisHub: Advantages

Technical:

- Sharing data over the network, but no firewall problems (HTTP / HTTPS)
- License free, open source, internet standards
- Platform independent, most basic client is a standard browser
- XML:
  - Data validation on upload (XML schemas)
  - Data transformation on request (XML stylesheets)
- Querying: SQL or XPath

Scientist:

- May modify there data provided as XML document at any time without corrupting the underlying database
- May dynamically add or delete search indexes, schemas and stylesheets

SeisHub: Disadvantages

Technical:

- Slower than “common” solutions
  - XML parsing during validation and indexing
  - Data overhead (XML verbosity)
- Infrastructure

Scientist:

- Seismologist != IT nerds

Table of Contents

ObsPy

- Python toolbox for seismologists
- Goal: facilitate rapid application development for seismology
- Modular extensible architecture
  - Waveform data: GSE2/GSE1, MiniSEED, SAC, SEG-Y, SH Q/ASCII, SEISAN, IRIS TSPAIR & SLIST, WAV
  - Inventory data: Dataless SEED, XML-SEED, RESP
  - Data request clients: ArcLink/WebDC, IRIS DHI/Fissures, SeisHub, IRIS Web Service, NERIES Web Service
  - Signal processing: Filters, triggers, instrument correction, rotation, array analysis, beamforming
  - Plotting: spectrograms, beachballs and waveforms
  - Waveform indexer

ObsPy

- Open source (LGPL or GPL)
- 6 core developers
- Platform independent (Win, Mac, Linux) and tested
- Test-driven development (TDD), currently ~500 unit tests (http://tests.obspy.org)
- Reliance on well-known third-party libraries (numpy, scipy, matplotlib)
- Reusing well established code, e.g. libmseed, GSE UTI
- Automated build of API documentation from source
- Binary distributions: PyPI (http://pypi.python.org) & Windows Installer
- Source code & community webpage containing tutorials, installation instructions, ticket system, mailing lists

http://www.obspy.org

ObsPy: Demonstration

ObsPy: Master Plan

- Limited number of human resources & imbalance between number of researchers and software engineers
- "Our" solutions:
  - outsource the problem by growing our own community ;)
  - focus on young researchers/students - "dogfooding"
  - small entry barriers
  - make people feel responsible for modules
  - fast developer rights to interested prospects
  - proper infrastructure, e.g. communication, development, automated testing, documentation, LMU independent
- Goal: overall increasing of productivity and sustainability

Table of Contents

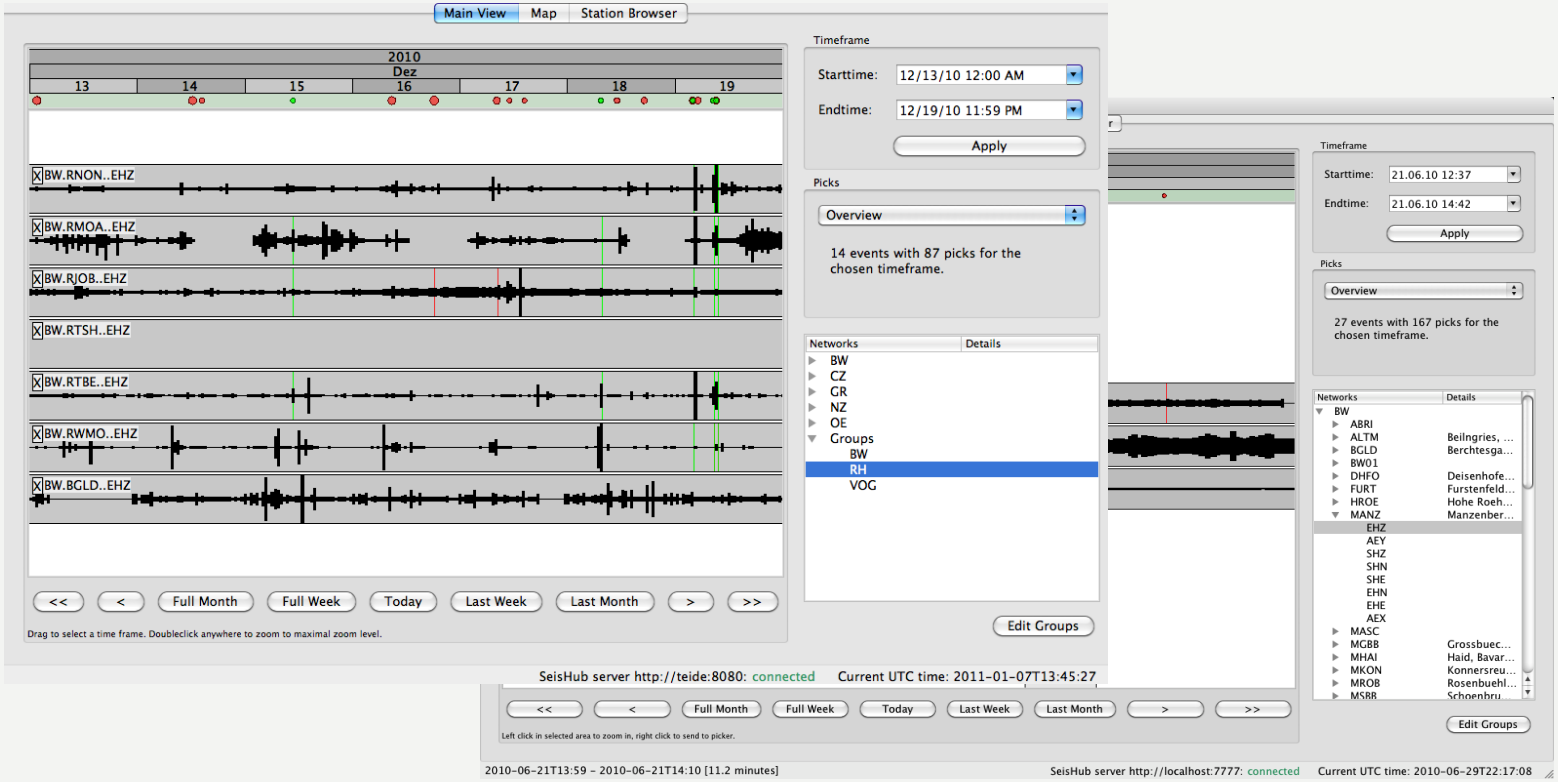
BayernNetz

- Seismological network of the Bavarian Seismological Service (Erdbebendienst Bayern)
- Designed for monitoring local seismic activity in Bavaria and bordering regions for events of $M_L >= 2.0$
- Data is collected, archived and analyzed at the Geophysical Observatory in Fürstenfeldbruck
- Currently 23 SP + 3 BB + 2 BB (GRSN) stations
- Waveform data received from BB stations are transmitted in real time to ORFEUS, GEOFON and BGR
- Neighboring networks may also access data directly

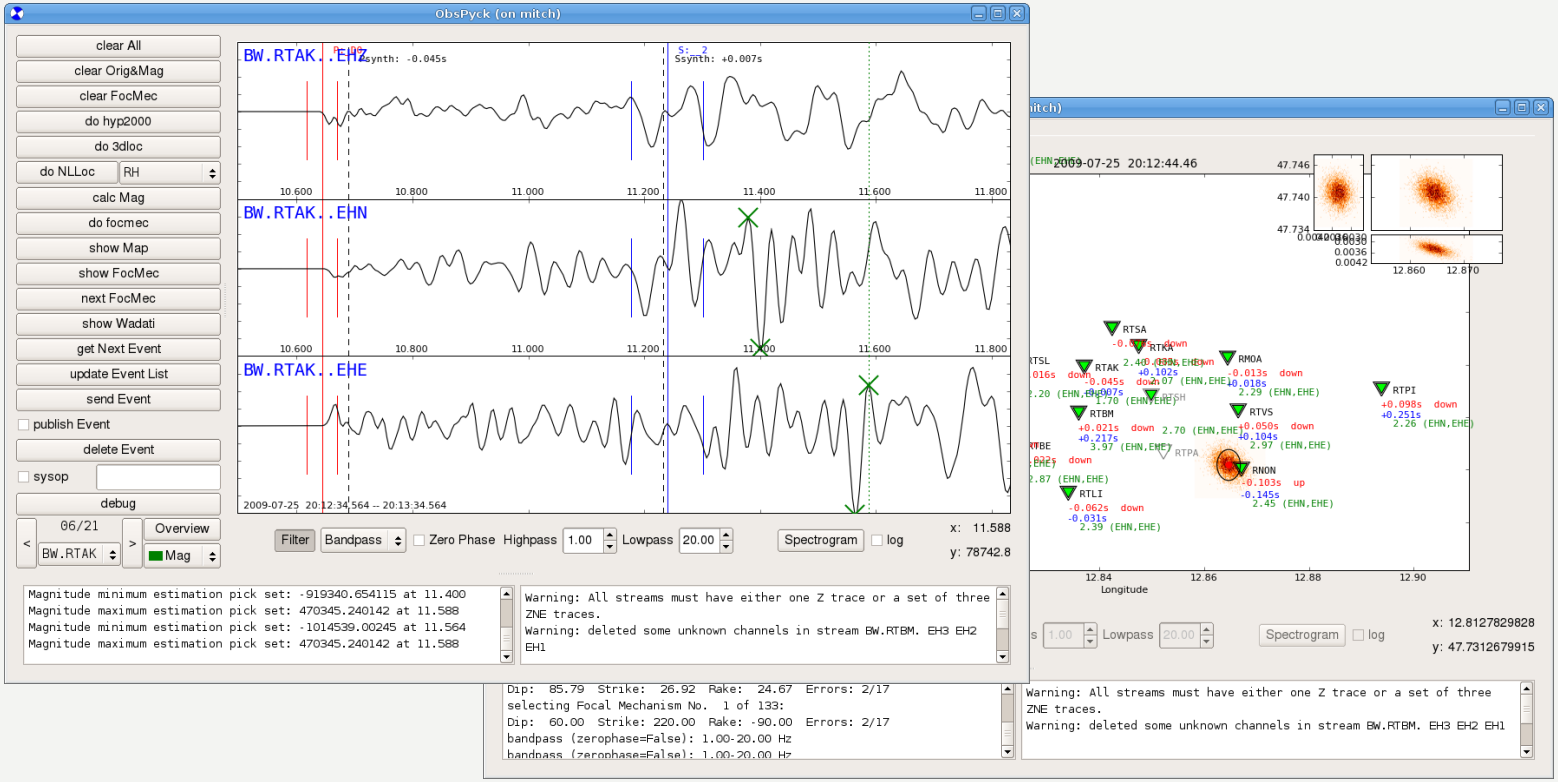
BayernNetz: SeisHub & ObsPy

- SeisHub
  - Replaced previously used GIANT database
  - Events retrieved from Earthworm (real time system) are automatically stored into SeisHub in a QuakeML-based format
  - Manual phase picking & event localization via ObsPyck (replacing PITSA)
  - Automated indexing of the MiniSEED waveform archive (SeedLink)
  - Stations available in XML-SEED
- ObsPy: seismological Python library and SeisHub client

SeisHub Viewer



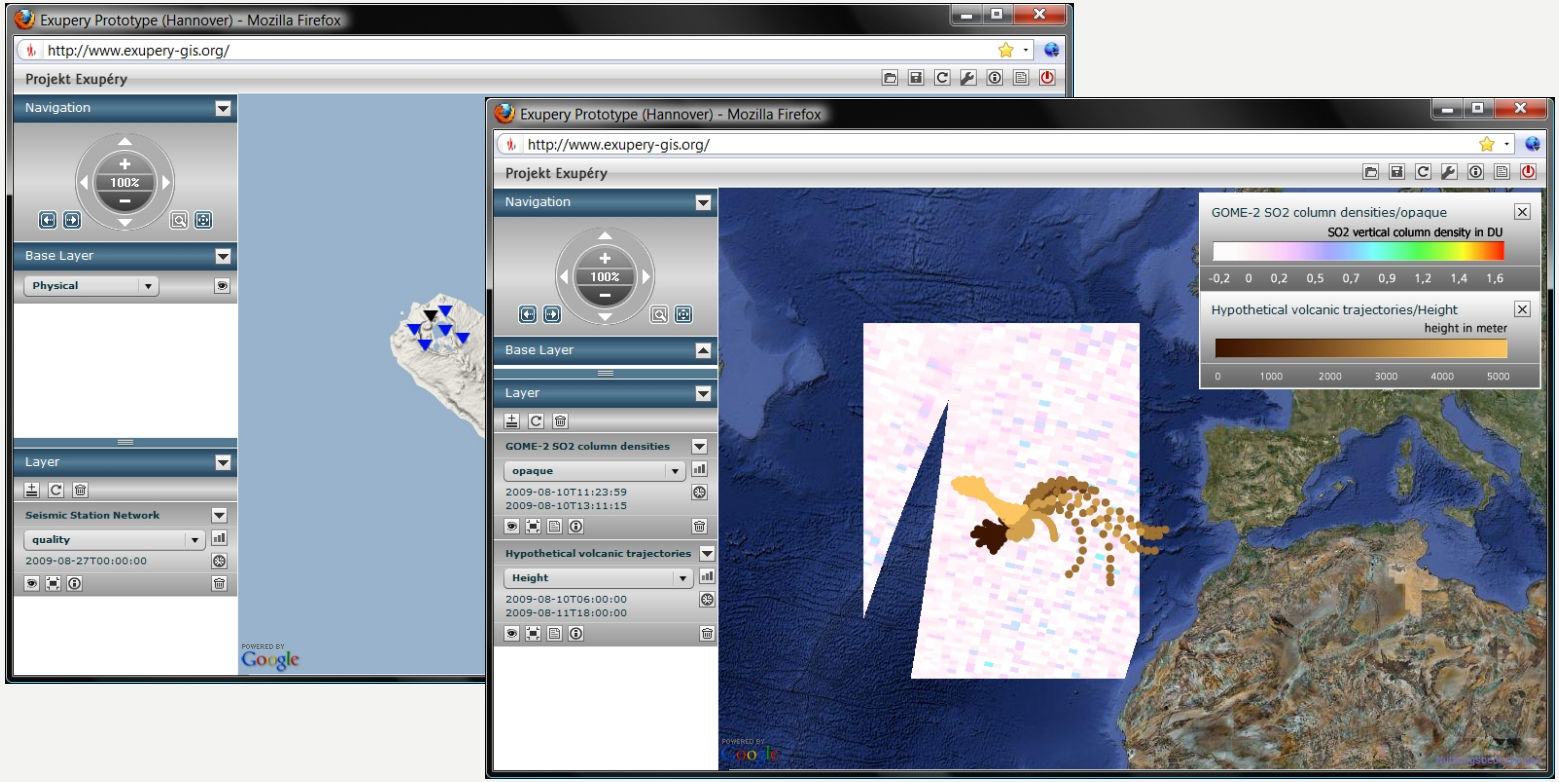
ObsPyck


ObsPyck (on mitch)
clear All
clear Orig&Mag
clear FocMec
do hyp2000
do 3dloc
do NLLoc
RH
calc Mag
do focmec
show Map
show FocMec
next FocMec
show Wadati
get Next Event
update Event List
send Event
publish Event
delete Event
sysop
debug
06/21
Overview
BW.RTAK
Mag
BW.RTAK..EHZ
BW.RTAK..EHN
BW.RTAK..EHE
2009-07-25 20:12:34.564 -- 20:13:34.564
Filter
Bandpass
Zero Phase
Highpass
1.00
Lowpass
20.00
Spectrogram
log
x: 11.588
y: 78742.8
Magnitude minimum estimation pick set: -919340.654115 at 11.400
Magnitude maximum estimation pick set: 470345.240142 at 11.588
Magnitude minimum estimation pick set: -1014539.00245 at 11.564
Magnitude maximum estimation pick set: 470345.240142 at 11.588
Warning: All streams must have either one Z trace or a set of three ZNE traces.
Warning: deleted some unknown channels in stream BW.RTBM. EH3 EH2 EH1
2009-07-25 20:12:44.46
Longitude
x: 12.8127829828
y: 47.7312679915
Dip: 85.79 Strike: 26.92 Rake: 24.67 Errors: 2/17
selecting Focal Mechanism No. 1 of 133:
Dip: 60.00 Strike: 220.00 Rake: -90.00 Errors: 2/17
bandpass (zerophase=False): 1.00-20.00 Hz

Exupéry VFRS

- Volcano Fast Response System (VFRS): mobile monitoring system in case of a volcanic crisis or volcanic unrest
- funded by the German Ministry for Education and Research (BMBF) within the "Geotechnologien" project
- Involves nine different research institutions allover Germany
- Primary goals
  - stable communication basis for stations in the field
  - expert system allowing scientists and local authorities to assess the data through a web-based GIS interface on top of a single centralized database
- Perfect demonstrator for the flexibility of SeisHub as geophysical database:
  - event-based and continuous data,
  - ground-based measurements and satellite data
  - time series (1D), images (2D), and models (3D)

Exupéry VFRS



Thank You!