SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

4.2: Report on metadata challenges and proposed solutions

Work package	4.1 . Expanding access to the European seismic monitoring infrastructure
Lead	John Clinton, ETHZ
Authors	John Clinton, ETHZ; Wayne Crawford, IPGP; Christos Evangelidis, NOA; Philipp Kaestli, ETHZ; Philippe Gueguen, ISTerre; Helle Pedersen, ISTerre; Javier Quinteros, GFZ; Jean-Marie Saurel, IPGP; Reinoud Sleeman, KNMI; Angelo Strollo, GFZ
Reviewers	[Name, Institution]
Approval	[Management Board]
Status	Final
Dissemination level	Public
Delivery deadline	31.10.2018
Submission date	31.10.2018
Intranet path	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

SERA deliverable 4.2 is a report on metadata challenges and proposes short term and long term solutions. There is particular focus on the short term solutions that can be implemented by seismic monitoring and archival infrastructures during the duration of the SERA project, that can lead to richer and better archives across Europe.

D4.2 was originally due on Month 12 (April 2018), the delivery has been extended to Month 18 (October 2018). D4.2 is directly related to Task 4.2, and partly related to Task 4.3. It is a prerequisite for beginning work on SERA deliverables 4.3 (Best practices guidelines, M24) and 4.4 (EIDA Metadata model standards, M36). At this stage, we expect that we request D4.3 and D4.4 to be jointly delivered in M30.

In addition to being a standalone report for SERA and a document that builds towards the wider SERA WP4 deliverables, we hope this deliverable will be used as an independent community document where the proposed solutions are implemented at European datacenters, and the proposed standards are candidates to become community standards. Community acceptance of the proposed solutions should go hand-in-hand with further SERA technical work (technical work cannot wait for full community acceptance) and be compliant with EPOS TCS-ICS services. At the same time we recognise the importance of global coordination of work and discussions towards international standardization of future formats in seismology and beyond.

1 Introduction

In order to build effective and sustainable services, the organization of data and metadata parameters should follow well-described standards. Standards for data and metadata are increasingly important in the geosciences. In seismology, standards and conventions have been agreed by the global community for some decades, and have coincided with (and been partly responsible for) a period in which the volume, accessibility and usage of datasets have exploded. Data sharing is now routine, and, compared to decades ago, conducting science is relatively unhindered by data management and conversion.

Building popular standards is not a trivial task, and requires community engagement and community agreement. Other geoscience communities without existing community standards that collect basically similar data and metadata often look to adopt seismological standards.

In seismology, standards are agreed within the auspices of the Federation of Digital Seismic Networks (FDSN, www.fdsn.org). Typically, members of the community propose standards, alongside documentation and best practices, to the FDSN, and the FDSN votes on their adoption. Nevertheless, once a standard is accepted, communities must invest in ensuring effective usability before it achieves widespread adoption.

Current standards in seismology are:

- miniSEED2.0 as the data format for evenly spaced raw seismological waveform timeseries (ref: http://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf)
- StationXML to describe metadata for seismic stations (ref: https://www.fdsn.org/xml/station/) (this supersedes datalessSEED (ref: http://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf)
- quakeMl1.2 to describe earthquake information (www.quakeml.org)

- FDSN webservices (event, station, dataselect) for discovery and dissemination of seismic data and earthquake information (www.fdsn.org/fdsnws)
- EIDA webservices (wfcatalog, routing) for discovery of waveform metadata and services (currently on the European level)

Though these standards are now widespread across and even beyond the seismology community, they have been developed for a specific seismology community and in many cases are not sufficient to accurately describe the datasets collected by communities that would like to adopt these standards.

This report intends to:

- describe challenges for integrating datasets in current seismological standards based on miniSEED / stationXML / quakeMl
- propose simple solutions in archiving new and existing datasets that allow the community to move forward fast, within the general confines of the existing standards
- indicate challenges that should be addressed but will require longer term efforts including changes or extensions to current standard data models
- support the increase in the variety of data types and the overall total volume of data archived in EIDA in a sustainable and effective manner.

The report consists of a number of Chapters for each of the communities outlined in Table 1. Each Chapter is prepared with contributions from experts in each field. The format for each Chapter follows that in the Appendix

The intention is to adopt these proposed standards / best practice guidelines within European datacenters / data collections. It is expected that these proposed standards will also be distributed directly to individual communities for further review, with invitation to comment, and may become candidates for future FDSN standards.

This procedure can also be followed by other communities who wish to propose changes to standards or adoption of best practice / conventions.

The list of data types addressed in the subsequent Chapters are in Table 1 and Table 2. Table 1 focuses on existing technologies that are using current seismological standards but would strongly benefit from 1) agreed and widespread conventions for addressing domain specific issues; and 2) extensions in the standard formats to incorporate rich collections of domain-specific metadata. Table 2 includes new monitoring technologies that will require careful community consensus, possibly including completely new formats, before they can become routinely integrated in seismic archives. For these new communities we provide only an overview listing requirements and new challenges to be addressed in the near future.

Data Type	**Requirements:**
Structural Monitoring	- Conventions for SEED channel naming. - Extended metadata to describe structure and sensor locations within structures **Community Acceptance:** COSMOS, ORFEUS Strong Motion
Ocean Bottom Seismometers (OBS)	- Conventions to address drifts in timing - Uncertain location, may drift over time - Additional information concerning deployment - Availability of raw and corrected waveforms (including description of

	the correction applied) **Community Acceptance:** OBS pools
QuakeMl Basic event description	- suport extending and flexible event types - managing catalogues **Community Acceptance:** FDSN, EMSC, ISC, NEIC
Volcano and NFO geophysical datasets	- Conventions for SEED channel naming. - Minor stationXML fix. - Easy to use metadata editor tool. **Community Acceptance**: EPOS NFO, Volcano
Moving Sensors (Slopes / Glaciers / Volcanoes)	- Conventions for how to handle moving sensor coordinates
Old strong motion records without precise timing	- Convection on SEED channel naming - Extended metadata to describe absolute or relative timing **Community Acceptance**: EPOS NFO, Volcano

Table 1: Existing Communities

Data Type	**Requirements:**
Large N	See the document on the next generation of miniSEED: https://docs.google.com/document/d/1ymAe9v1rUuucpY7ai5ilKsD7V1ejwt6GxQQmJ5IevDI/edit?usp=sharing **Community Acceptance**: FDSN
Distributed Acoustic Sensing (DAS), or interrogated fibre optic cables	Preservation of raw data as well as possibility to convert in FDSN standard formats. Sensor/acquisition description may not fit the actual model provided in station_xml. Need to explore on the fly conversion tools or decide a strategy about virtual sensors along the fibre. **Community Acceptance**: FDSN
Cheap sensors including community networks, schools / mobile phone sensors -	
Volcano and NFO geochemical datasets	- New stream identifier scheme needed - Precisions on the sampling condition needs to be added to metadata - Question about a different format than for manual gas and water sampling and subsequent analysis **Community Acceptance**: EPOS NFO, Volcano

Table 2: New Communities

A significant number of deficiencies highlighted here are also addressed in a recent FDSN-moderated discussion on creation of a new data format for seismology (https://docs.google.com/document/d/1ymAe9v1rUuucpY7ai5ilKsD7V1ejwt6GxQQmJ5IevDI/edit?usp=sharing). This effort identifies that the most pressing deficiencies in miniSEED arise from the inability to handle data from new experiments with very large number of sensors. The usage of the Station-Network-Channel-Location (SNCL) identification for time series must be significantly expanded, or even redesigned, to accommodate for a variety of anticipated sensor deployment configurations with an increased number of components. A revision of the stream identification is required in order to support this, i.e. large number of sensors, dense sensor spacing, different sensor types, distinguish between raw, processed and synthetic data, and refer to a full description of the processing (provenance). The current identification cannot handle this:

- Network Code: currently limited to 2 characters and is too limited. Conventions, not format standards, are being used to work within the current miniSEED
- Station Code: as the number of stations will increase in deployments, the number of characters (currently 5) needed to identify time series should increase to accommodate larger numbers of stations (of the same order as used in the exploration industry in a rational manner).
- Location Code: the current two characters are not adequate to represent arrays of sensors using a reasonable naming convention. The next generation of the format should allow up to one million sensors deployed in arrays using the best estimates of sensor growth in the next one or two decades.
- Channel Code: as the types of sensors increase the method within current miniSEED to provide unique sensor identifiers has been exhausted. Furthermore, the need for identification of derivative and synthetic time series is increasing.

The eventual adoption of a next generation miniSEED will provide basic support for many of the long-term solutions proposed in this document.

2 Template for Domain Reports

2.1 Data Summary and Target Community

- Very brief summary of scope of data collected, including current approaches, key stakeholders (data collectors, data users), how data is currently disseminated

2.2 Current challenges / shortcomings

- List of metadata challenges that are directly addressed in following short-term / long-term sections
- Can include issues beyond metadata, such as dissemination

2.3 Short term solutions

- Solutions within confines of existing data format

2.4 Long term solutions

- Solutions that would require changes or extensions to data format standards

2.5 Integration in EIDA

- Includes outlook for dissemination services

3 Appendix A: Structural Monitoring[1]

main author: John Clinton / community expert: Philippe Gueguen

3.1 Data Summary and Target Community

It is increasingly common to deploy high quality seismic sensors across structures, often in large numbers, often with continuous recordings. The scientific motivation for such monitoring includes: characterise the dynamic behaviour of the structure; diagnose and understand structure damage during strong motions by recording the response to earthquakes; and characterise ageing Structures monitored include buildings (particularly high rise), dams, bridges, towers and tunnels. These fields are often referred to as Structural Health Monitoring or Condition Based Maintenance in the domain community.

It is becoming common that long term structural monitoring is performed by teams that operate seismic networks in general, and using equipment that is conversant with seismic network standards, so it is sensible to use try to include these datasets into the existing archives and use the same data / metadata formats.

The majority of users of this data are engineers and engineering seismologists. In general, they are not familiar with seismological data access and processing, and not willing to learn them. An effective dissemination tool should provide access in formats usable to engineers, eg text or matlab files for waveforms, that may also combine the metadata. Various legacy formats for this exist.

On key community issue is the data policy. Many monitoring infrastructures continue to be operated by private agencies who do not observe standards or share the data. A strong motivation of defining standards and demonstrating their success is also to motivate private networks to share their data via the seismological data centers. This will also have benefits for the quality of the data and long-term sustainability.

3.2 Current challenges / shortcomings

List of challenges and shortcomings:

1. Lack of agreed guidelines for data archival and metadata formats
    -> propose to use existing seismological standards for short term and long term solution
2. When seismological standards are used, there is no standard convention / best practice
    -> short term solution addresses this
3. Additional and important domain specific information is missing (eg relative sensor positions, even basic description of the building)
    -> do not abuse existing but underused stationXML options such as Vault or StationDescription
    -> long term solution proposes using quakeML2.0 extensions
4. End-user specific dissemination tools are missing
    -> long term solution required

[1] Many ideas presented here were originally collected within an effort in EPOS PP to plan future archives and services for the Structural Monitoring community. The authors recognise and thank Erdal Safak, the Task Leader

We recommend that the structural monitoring community use existing seismological standards for data storage, metadata management and data distribution, with appropriate extensions. The data should be labelled according to FDSN standards (Seed naming convention, stationXML metadata) and should be archived and distributed using the EIDA infrastructure. **However, instrumentation in structures is specific and the arrays cannot be sufficiently described using the existing stationXML, in particular relating to basic information about the structure and foundations; the position of the sensors inside the structure (eg for a building, whether in NE corner, attached to a beam / column / floor diaphragm), and support for high-precision, but often only relative locations of the sensors.** Additionally, though there are well- described naming conventions, there is not a consistent application of the standard by managers maintaining structural monitoring data.

Since some members of the community have been recording data from buildings for decades, there are a number of existing well-instrumented structures that have taken different approaches in managing their datasets. Over the past few years, many structural monitoring arrays have already been added to seismic archives at eg IRIS or RESIF. A number of different approaches to labelling the array data have been taken. Each of these follows the SEED convention, where codes are defined for the Network, Station, Location and Channel. We outline 3 examples in the final section of this Appendix, from 3 different buildings, 1 in the US and 2 in France.

The best practice proposed from here takes into account the experience gained in these case studies.

3.3 Short term solutions

Metadata Modifications

We use as basic starting point the existing Seed standard for channel naming, and stationXML for metadata information.

To more fully characterise the structure and the network inside the building, extensions to the metadata model are required, and these are described in the Long Term section. For both channel naming conventions and metadata, we provide suggestions for best practice.

Conventions within Existing StationXML

Where relevant we follow the Seed naming convention – see Appendix A of https://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf

The following highlights key fields in StationXML that should be followed for each monitored structure.

Network.Code (2 Characters) : use the network code of the seismic network that installs and manages the data, eg RA for French RAP / Resif ; CH for Swiss Seismic Network. Must be registered with FDSN http://www.fdsn.org/networks/

<u>Station Information</u>

Station.Code: (2-5 characters) name of the structural array being monitored. Use a single station name for the entire array, eg for a single building, tower, bridge, tunnel or dam[2]

[2] In this convention, the station code signifies a monitoring array and is not restricted to be a single instrument or data-logger. Multiple sensors can be included at a single station, and are distinguished by the channel name and location codes. Additional guidelines you may choose to follow are: the station code can be 1-5 alphanumeric characters, but use at least 2. It is not case sensitive and is traditionally written in upper case. Most networks use combinations of letters and numbers that are temporary networks, use numeric station codes. In those cases numbers shorter than five digits are NOT zero-filled.

Station.latitude and Station.longitude: latitude and longitude in degrees of the reference point of the structure (arbitrarily chosen). WGS84 is assumed unless otherwise stated. Should have 10cm resolution (6dp in degrees).

Station.elevation[3]: elevation of a reference point for the array (arbitrarily chosen), in m. Suggested convention is to use
Building, Tower: ground level at a given point beside the building
Dam: could be mid-point of dam crest, or base of dam
Bridge, Tunnel: could be surface level at one entry point

Station.depth[4]: depth from the free surface (or where it would be in absence of the structure) to the station.elevation. Positive is downward.

<u>Location Information</u>

Channel.LocationCode (2 characters): in a structural array, the location code is used to distinguish between sensors at different locations in the array. The location code can be opaque (is not intended to be interpreted), but can also provide some information about the sensors. Possible non-opaque conventions can be:

Option 1: ideal for vertical arrays with few sensor elements (buildings, boreholes)
1st character indicates the vertical order, in ascending order, from basement to roof (buildings)
2nd character indicates the position along the structure, eg W for west side boreholes:
SF for surface
BT for bottom
Intermediate sensors can use actual depths in m if less than 100m, otherwise can use M[1,2,3] for intermediate depths in order of depth from surface.

Option 2: ideal for vertical arrays with many sensor elements (buildings, boreholes)
1st character in descending order (1 to Z) according to the position along the structure. For example from building top roof to the bottom of borehole in case of building+borehole array.
2nd character : 0: middle (0 by default for borehole) - 1 to 8 in clockwise : N, NE, E, SE, S, SW, W, NW position.

Option 3: ideal for horizontal arrays (bridges, tunnels, dams)
1st character : 0 for deck level, or top dam level, and in descending order following the geometry of the structure.
2nd character: in descending order from the distance to the reference point of the structure. Changing according to the structure.

Note: station xml refers to seed for station codes, and seed is unaware of these conventions. Thus you may find current (and future), valid data streams with these location codes, but different meaning. To fix this, a new version of stationXML is required.

For example, station "334" would NOT be written as "00334". Station codes can be registered at ISC http://www.isc.ac.uk/registries/registration/

[3] Note that in stations with one sensor, the elevation must be the elevation of the sensor respectively, because this is how "elevation" is defined in stationXML

[4] As 2 for depth instead of elevation

Channel Information:

Channel.Code (3 characters): for structural arrays, we assume we are collecting (triaxial) accelerometric data, so channel name should be

- HN[Z23] for orthogonal sensors with horizontal sensors oriented in the principal directions of the structure (standard practice in structural monitoring),
- HN[ZNE] for orthogonal sensors oriented in traditional components with respect to North.

The azimuth of the sensor to North is recorded in the stationXML. If uniaxial sensors are used, follow the same convention with HG[ZEN23] where appropriate. Note the azimuth can change from sensor to sensor, and the individual dip and azimuth for each sensor are defined in the metadata.

Channel.latitude and Channel.longitude: latitude and longitude in degrees of the actual sensor. Should have 10cm resolution (6dp in degrees). WGS84 is assumed unless otherwise stated. It is accepted that this may suggest an absolute location precision that is, in reality, a level of precision only relative to the station.latitude/longitude another position in or near the structure.
Note: StationXML does not support relative position to station latitude / longitude

Channel.elevation: the elevation of the sensor, m

Channel.depth: depth from the free surface (or where it would be in absence of the structure) to the station.elevation. Positive is downward.

Channel.azimuth: azimuth of the channel orientation w.r.t. true North (Z=0°; N=0°; E=90°)

Channel.dip: dip of the channel orientation w.r.t vertical (if ZNE: Z=0°; N=90°; E90°; if Z23: Z=0°; 2=90°; 3=90°)

3.4 Long term solutions

I. Modifications to StationXML: Option in Station XML to provide distances for sensors relative to the station instead of absolute values.

II. Extensions to basic StationXML / QuakeML

- Site description Adopt QuakeML2.0 site characterisation package for free-field station with extension (Class Description, Site Morphology, Vs30...) + literature source
- Structure description Review and extend QuakeML2.0 station characterisation package. Conventions to be used for the following parameters:

Station housing.Class: Describes the basic type of structure.
Requires basic fixed vocabulary: Building, Dam, Bridge, Tunnel.
If building, append correct EMS classification

M1: rubble stone;
M2: adobe;
M3: Simple stone;
M4: Massive stone;
M5: Unreinforced masonry;
M6: unreinforced masonry with RC floors;
M7: Reinforced or confined masonry;

RC1: RC frame without Earthquake Resistant Design (ERD);
RC2: RC frame with moderate ERD;
RC3: RC frame with high ERD;
RC4: RC shear walls w/o ERD;
RC5: RC shear walls w/ moderate ERD;
RC6: RC shear walls w/ ERD;
S: steel structures;
W: wooden structures. STRING

Station housing.Description:
For buildings, use EMS98 classification:

Station storey.Count: for Buildings only. Indicates the number of stories including roof and basement levels. In format TotalNumerofLevels.numberofBasements, eg 58.2 indicates 58 floors above ground and 2 basement levels. Usually, the number of stories must include the ground floor level and be the number of main floors above ground, including any significant mezzanine floors and major mechanical plant floors. The number of stories is then given as the number of horizontal stiff diaphragms and then mechanical mezzanines or penthouses must not be included if they have a significantly smaller (2/3) floor area than the major floors below.

Foundation: Type of foundation

III. Additional Information on the structure
A more complete description of the structure can be prepared, including photos / figures / drawings. It may also contain more detailed information about the instrumentation. The content can be compiled in a PDF format and distributed through a portal. The goal is to provide further context, but it is not expected this will be sufficiently detailed to develop sophisticated numerical models for the structure. They may though suffice for 1) simple derivation of models in combination with analysis of the observed response of the structure to seismic loading - as required for performance based seismic design or prediction of structural damage; and 2) identification of attributes that impact the seismic vulnerability as proposed by the standard vulnerability scales (EMS98, RiskUE, HAZUS, GEM Taxonomy).

For **Buildings**, we propose to use the GEM Building taxonomy - see references: https://www.nexus.globalquakemodel.org/gem-building-taxonomy/overview. This includes:
1. Direction – the orientation of building(s) with different lateral load-resisting systems in two principal horizontal directions of the building plan which are perpendicular to one another
2. Material of the lateral load-resisting system - e.g. "masonry" or "wood"
3. Lateral load-resisting system - the structural system that provides resistance against horizontal earthquake forces through vertical and horizontal components, e.g. "wall", "moment frame", etc.
4. Height - building height above ground in terms of the number of storeys (e.g. a building is 3-storey high); this attribute also includes information on the number of basements (if present) and the ground slope
5. Date of construction or retrofit - the year in which the building construction or retrofit was completed. Can include multiple dates of occurred multiple times.
6. Occupancy - the type of activity (function) that the building is used for
7. Building position within a block - the position of a building within a block of buildings (e.g. a "detached building" is not attached to any other building, 'semi-detached' and 'row' indicate sharing of walls.)
8. Shape of the building plan - e.g. L-shape, rectangular shape, etc.

9. Structural irregularity - features of a building's structural arrangement that are irregular; such as one story is significantly higher than other stories, or the building has an irregular shape. Also the change of the structural system or materials that produce known vulnerability during an earthquake fall into this category. Re-entrant corner and soft storey are examples.
10. Exterior walls - material of exterior walls (building enclosure), e.g. "masonry", "glass", etc.
11. Roof - this attribute describes the roof shape, material of the roof covering, structural system supporting the roof covering, and the roof-wall connection. For example, the roof shape may be "pitched with gable ends", roof covering could be "tile", and the roof system may be "wooden roof structure with light infill or covering".
12. Floor - describes the floor material, floor system type, and floor-wall connection. For example, the floor material may be "concrete", and the floor system may be "cast in-place beamless reinforced concrete slab".
13. Foundation - that part of the construction where the base of the building meets the ground. The foundation transmits loads from the building to the underlying soil. For example, a shallow foundation supports walls and columns in a building for hard soil conditions, and a deep foundation needs to be provided for buildings located in soft soil areas.
14. Additional documents: plan, pictures etc...

For Bridges, we propose to use the terms based on the "Handbook for Bridge Inventory" produced by the Norwegian Public Roads Administration as part of the inventory module of the bridge management system (see reference: (https://www.tsp2.org/library-tsp2/uploads/48/Handbook_of_Bridge_Inspections_Part_1.pdf). This includes:

1. The bridge category - in relation with the kind of traffic the bridge is designed (examples: road bridge, pedestrian, railway bridge etc.)
2. The type of static system - given by how the superstructure acts when carrying loads (for example: simply supported system, continuous system, cantilever system, arch system, etc.)
3. The superstructure - the nature of the element that carried the traffic (for example: slab, beam, deck, etc.)
4. The special components for cable structures - for example, cables, hangers etc.
5. Additional information that provide relevant description of the bridge - for example, the substructures (piers, towers etc.), the interaction with the ground that may have influence on the structure (river course, embankments etc.)...

For Dams, we propose to use the description given by the British Dam Society (see reference: https://britishdams.org/educationcareers/about-dams/types-of-dam/). This includes:

1. Arch dams - made from concrete, curved in the shape of an arch, with the top of the arch pointing back into the water.
2. Buttress dams - made from concrete or masonry, with a watertight upstream side supported by triangular shaped walls, called buttresses.
3. Embankment dams - made from earthfill or rockfill.
4. Gravity dam - made from concrete or masonry, with gravity that holds it down to the ground stopping the water in the reservoir pushing it over.

3.5 Integration in EIDA

I. Archival and metadata

Minimum

- Waveform data archived with SNCL as proposed above
- StationXML created with conventions as indicated in Short Term Solutions.

Optional

- Populate the quakeML2.0 extension for site and station characterisation
- Prepare PDF with additional information / photos / figures

II. Distribution
Minimum

- Data discoverable and accessible using standard webservices / portal

Optional

- Structural monitoring specific portal to be developed for domain-specific distribution, that provides access information in StationXML / quakeML2.0 extensions as well as the PDF. Allows discovery of all structural arrays and supports event and continuous data downloads in formats consistent with domain expectations

3.6 Examples of Previous Implementations

Example 1 - Factor Building at UCLA, Los Angeles, USA - Building array as a specific network

The 17-story, steel-frame Factor Building in Los Angeles, USA houses the UCLA's Center for Health Sciences. The monitoring system is composed of 72 FBA-11 (from Kinemetrics Inc.) accelerometers, four horizontal channels at every floor above the ground level. The basement and sub-basement floors have two vertical and two horizontal channels each, as shown in the figure below. The horizontal sensors are oriented north-south and east-west principal directions of the building. The nine units of eight-channel recorders and digitizers are located at the top floor of the building, recording continuously at 200 sps.

A 100-m-deep borehole and a surface tri-axial accelerometers are also installed 25 m away from the building. Their recorder is synchronized with the recorders in the building.

See information: https://service.iris.edu/fdsnws/station/1/query?network=FA&level=cha&format=text
The Factor Building has a specific network code and the miniseed file naming convention is as follows:

Network Code FA (only for this array)
Station Code the station names are unique for each of the 14 digitisers in the array – ie the station name for each sensor is determined by the digitiser it is connected to.
In Building: FABA: Sub-basement and basement
FABB: 1st and 2nd floors
FABC: 3rd and 4th floors
FABD: 5th and 6th floors
FABE: 7th and 8th floors
FABF: 9th and 10th floors
FABG: 11th and 12th floors
FABH: 13th and 14th floors
FABI: 15th floor and roof
FABS: borehole
Free-field: FABW, FABX, FABY, FABZ
Channel Code (the building is oriented directly NS and EW)
HN(E,N,Z)=high-gain, 100 sps, EW, NS, or vertical
EN(E,N,Z)=500 sps
Location Code (single-component FBA11/Episensors)
1st character indicates height in building
Y=subbasement, X=basement
1-9= floors 1-9

A=floor 10, B=floor 11, C=floor 12, D=floor 13, E=floor 14, F=floor 15, G=roof
2nd character indicates position on floor
(N,S,E,W = corner of floor where FBA is located)
Exceptions:
DH=downhole (borehole)
UH=uphole (borehole)

Example 2 - Ophite Tower, France - Building array with a single multi-channel Digital Acquisition System, part of a national network

Since October 2008, Ophite tower, Lourdes (France) has a Kinemetrics FBA-11 accelerometer network composed of 2 or 4 horizontal channels per floor except for the basement and subbasement which have two vertical and two horizontal channels each corner. The horizontal sensors are oriented along the longitudinal and transverse direction of the building. The digitizer, located in the building's basement, consists of 24-channels Kephren of continuous data since 2011.

See information: PYTO station webservice

The PYTO network is part of the RA network and the miniseed file naming convention is as follows:

Network code RA (the network code for the RAP network that operate the array)
Station code PYTO
Channel code
HN(2,3,Z)=high-gain, 100 sps, longitudinal (N135°), transverse (N225°), or vertical
Location code (single-component FBA11/Episensors)
00=roof (floor 20) - corner NE
01=roof (floor 20) - corner SW
02=floor 17 - center
03=floor 14 - center
04=floor 10 - center
05=floor 10 - corner SE
06=floor 06 - center
07=floor 02 - center
08=basement (floor 00) - corner NE
09=basement (floor 00) - corner SW
10=basement (floor 00) - corner NW

Example 3 - Grenoble City-Hall - Building array with several multi-channel Digital Acquisition Systems, part of a national network

Since November 2004, Grenoble City-Hall (France) has a Kinemetrics FBA-11 accelerometer network composed of 6 3C Episensors at the top and the basement floor (Figure 3). The horizontal sensors are oriented along the longitudinal and transverse direction of the building. The digitizers, located in the building's floor right to the sensor, consist of 6 MiniTitanXT, each recording 3 channels of continuous data since 2010.

See information: GCH station webservice

The GCH network is part of the RA network and the miniseed file naming convention is as follows:

Network code RA
Station code
OGH1: basement (floor 00) corner SW

OGH2: basement (floor 00) corner NE
OGH3: basement (floor 00) corner SE
OGH4: floor 13 corner SW
OGH5: floor 13 corner NE
OGH6: floor 13 corner SE
Channel code
HN(2,3,Z)=high-gain, 100 sps, longitudinal (N330°), transverse (N60°), or vertical
Location code (single-component FBA11/Episensors)
00=for all stations

4 Appendix B: Ocean Bottom Seismometers[5]

main author: Angelo Strollo / community expert: Wayne Crawford

4.1 Data Summary and Target Community

Ocean bottom seismometers (OBS) provide seismological access to the 70% of the earth's surface that is covered by water. In Europe a relevant number of OBS parks are providing instruments for temporary campaigns, among them: INSU-IPGP and Geoazur (France); AWI-DEPAS, Geomar, UHamburg and BGR (Germany); IGPAS (Poland); ROA and UTM-CSIC (Spain). With the available parks several OBS experiments are carried out around the world, the latest coordinated effort between French and German groups has been carried out in the context of the AlpArray project with data being archived in EIDA. In most cases, in particular within the AlpArray community, users are requiring seamless integration of OBS data with land stations in order to easily access and process them. Despite the recent efforts still a large number of data sets are saved in FDSN-compatible seismological databases. The unsaved data often has incomplete metadata and is therefore hard to recover later on. The distribution of data by OBS facilities is hampered by a lack of clear standards and procedures for archiving data and metadata. In order to facilitate data ingestions in the EIDA archives W. Crawford in coordination with the main OBS parks in Europe proposed the following actions that are integrally available in a document submitted to the FDSN and partially summarized in the following basically divided in three parts:
1) OBS-specific standards and "best practices" for using the stationXML and miniSEED formats to make OBS data the most clear and easy to use;
2) post-processing tools for OBS data that should be made available to seismologists in order to reduce OBS-specific problems and take advantage of OBS-specific data possibilities.
3) modifications to the stationXML and miniSEED formats that will allow OBS (and other) data to be better informed;

4.2.Current challenges / shortcomings

Observations at the Ocean floor are mostly similar to the on land observations but with a number of complications due to the environmental conditions in which instruments are deployed and operated, in particular:

Recording
- Non-standard loggers (required in order to minimize volume and power consumption,
- completely autonomous in energy, operation and time base)
- Proprietary data formats
- Clock is synchronized only at beginning and end of deployments

Sensors
- Horizontal seismic channels generally not geographically oriented
- Pressure sensors are also included
- OBS sensors themselves may drift and move over the course of a deployment

Noise
- Seafloor currents

[5] Guidelines presented here are derived from the "OBS data/metadata proposal" by Wayne Crawford (Version: 201809) initially submitted to FDSN Working Group V Portable Instrumentation in November 2016 (http://www.fdsn.org/message-center/thread/471/)

· Motion under ocean waves
· Strong sea-surface reflections

Given the peculiarities listed above in order for Ocean Bottom Seismometer (OBS) data to be as useful as possible, OBS-specific properties (clock drift, horizontal channel nomenclature, water depth and location methods/precision) should be included in the data and metadata in the same way, regardless of what OBS facility collected the data. Some standard OBS data-specific software would also be very useful for helping scientists to get the most of this data.

Specific challenges in OBS data curation

1: Preparing data and metadata
· Standard tools such as NRL do not have OBS data logger information. Most OBS parks are not ready to create NRL libraries/templates and NRL libraries (based on RESP files) do not completely inform StationXML.
· For each deployment, same instrument must be combined with new location, station and network information.
· Clock drift must be documented (and corrected).
· Standard channel names must be defined and documented
- Horizontals = « {H,L}1,2 »
    § GSN definition, left-handed: geometrically « 1 » corresponds to « N » and « 2 » to « E »
    § Others say « 2 » and « 3 » should be used for horizontals, but IRIS DC is full of horizontal OBS channels named « 1 » and « 2 »
- Pressure:
    § « DH » (hydrophones)
    § « DO » (absolute pressure gauge)
    § « DH » or « DF » ? (differential pressure gauge)

2: Making the data as useful as possible
· Reorienting the horizontal sensors
· Verifying clock corrections
· Removing current and ocean wave noise

3: Documenting any known movement of the sensor
· refer and follow guidelines set in Appendix G - Moving sensors

4.3. Short term solutions

The following "best practices" are intended to achieve the best possible description of the data within the current standard data and metadata formats to facilitate the data usage also outside the OBS specific domain. Further additions/extensions to the current standards are provided in the next section.

<u>Timing corrections</u>

OBS clocks generally have a non-negligible drift because of the lack of GPS signal at the seafloor. The resulting time offsets must be corrected or at least indicated in any data archived at data centers. OBS time bases are generally chosen to have small and first degree linear drift. Their drift is calculated by synchronizing the instrument clock to GPS before the deployment and then comparing the instrument clock to GPS after the deployment. If the instrument clock cannot be compared to GPS at the end of the experiment, the drift can be calculated a posteriori by calculating the noise correlation between this instrument and another synchronized instrument over the length of the experiment. Information about the existence of linear clock drift, its value if measured and its probable range if not measured, should be provided in the data and metadata. We recommend the following practices:

StationXML

Indicate the timing correction in <Comment> or <CommentList> fields, as follows:

```
<CommentList>
    <Subject>Linear Clock Correction</Subject>
    <List>
        <Value>“time_base: Seascan MCXO, ~1e-8 nominal drift"</Value>
        <Value>“reference: GPS"</Value>
        <Value>“start_sync_reference: 2015-04-22T09:21:00Z"</Value>
        <Value>“start_sync_instrument: 0"</Value>
        <Value>“end_sync_reference: 2016-05-28T22:59:00.1843Z"</Value>
        <Value>“end_sync_instrument: 2016-05-28T22:59:02Z"</Value>
    </List>
</CommentList>
```

If the <CommentList> modification is not accepted as an addition to stationXML, bundle the same in a <Comment>, using JSON syntax:

```
<Comment>
    <Value>“{Linear Clock Correction: {time_base: Seascan MCXO, ~1e-8 nominal
drift, reference: GPS, start_sync_reference: 2015-04-
22T09:21:00Z,start_sync_instrument:  0, end_sync_reference: 2016-05-
28T22:59:00.1843Z,end_sync_instrument: 2016-05-28T22:59:02Z}}”</Value>
</Comment>
```

Absolute dates are used because they are unambiguous. “drift” or “slew” values are derived values and there is no standard for whether a positive value means the instrument is faster than GPS or vice versa.

miniSEED

Because of the lack of international agreement over whether to provide datacenters with clock-corrected or “raw” data, currently both can be provided: the “raw” data have the data quality field = “D” whereas the corrected data have data quality = “Q”. For clock-corrected data there three main possibilities under discussion within the community: a) Indicate the time correction in each record header but do not apply it (RAW); b) Indicate the time correction in each record header and apply it (SHIFTED); c) Resample the data at the originally intended rate (RESAMPLED). With option b (SHIFTED) being the preferred option as it allows the user to work with time corrected data which has not been modified but for which the time is as close as possible to GPS time. Until consensus is reached, there is the need to distinguish between these methods. If the time correction has been calculated:

Orientation information

- StationXML

Add new <Azimuth> (with plus and minus errors) for horizontal levels in StationXML, with a channel-level comment indicating how the orientation was determined.

Additional issues/fields:

- StationXML

Use Station <CreationDate> and <TerminationDate> fields to specify when the data was supposed to start and end, and <StartDate> and <EndDate> to specify when it actually starts and ends. This is best practice for temporary deployments in general, not only for OBS.

Within each Channel, set <Type>CONTINUOUS</Type> and <Type>GEOPHYSICAL</Type>

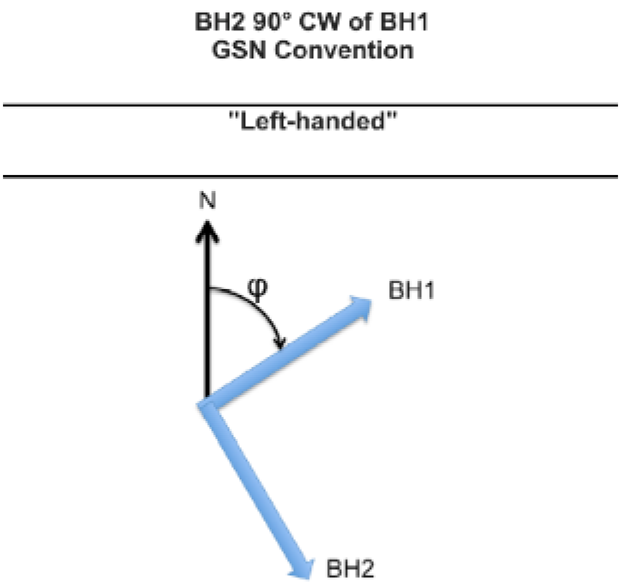
- miniSEED

Channel naming –
Pressure channel names
· Name hydrophone channels ?DH (“hydrophone”).
· Name differential pressure gauge channels “?DF” (“infrasound”)?[6]
· Name absolute pressure gauge channels “?DO” (“outside”) [7](OO uses “?DO”)

Channel Orientation Codes
If the instrument was not oriented along the tradition axes (N-S and E-W), the orientation codes for the horizontal channels should be “1” and “2” according to the geometry specified by the GSN standard shown below. “Azimuth” should be set to 0 for the “1” channel and 90 for the “2” channel. Uncertainties for both should be set to 180. Below an example how to specify Azimuths for the horizontal channels:

BH1: <Azimuth minusError="180.0" plusError="180.0" unit="DEGREES">0.0</Azimuth>
BH2: <Azimuth minusError="180.0" plusError="180.0" unit="DEGREES”>90.0</Azimuth>

This allows automatic orientation codes to get the proper geometry between the “1” and “2” channels, while indicating that the geographic orientation is unknown.



[6] Current (IRIS) practice is to name DPG channels ?DH

[7] Corresponds to OO naming convention (verify for IRIS: Cascadia Expt). An alternative would be ‘?TZ’ (tide gauge), though the “pressure” aspect is good to specify

Site names for repeated deployments
If OBSs are deployed repeatedly at one site (to make a long series), use an incrementing alphanumeric character at the end of the station name, to indicate subsequent deployments (i.e., A01A, then A01B then A01C for subsequent deployments at the same approximate site). Enter your logger and analog filter information into the NRL.

5.4. Long term solutions

In the following a few modifications to the stationXML and miniSEED formats are proposed aiming at better representation of OBS (and other) data. Although not described in this document some software exists for the post processing of OBS data. Long term needs of this community should include the development of a software toolbox for OBS data and metadata preparation and validation.

Station_xml

Add a “Water level” field. This is useful for removing/exploiting water surface reflections. In general, this would be set to 0 (sea level), but would be different if deployments are made in lakes or water-filled boreholes We chose “water level” rather than “water depth” because the default value would be “0” rather than “-elevation”.

This text field would indicate how the sensor position was determined. Probably often “GPS”, but could also be something like “Laser-based distance and compass-based angle from location 00” or “Acoustic survey”, etc.
Option: Add a “Measurement_method” attribute to uncertainty Double. That would allow one to also specify, for example, how Azimuth and Dip were determined.

Add a “CommentList” type. Would allow several related comments to be grouped together. Similar to the <Comment> type except that a <Subject> field would be added and <Value> would be changed to <Values> or <List> with multiple strings allowed.

Allow versioning. Some mechanism for specifying the version (perhaps with a means of specifying changes between versions). This is a general improvement for all communities, not particular to OBS.

miniSEED

Allow sampling rate to be specified as double precision. This is the only way to accurately represent OBS clock rates, which are regular but off of the specified sampling rate by a factor of approximately 1e-8 (MCXOs) or 1e-9.5 (CSACs), requiring 27- or 32-bit floating-point mantissas, respectively, to be correctly specified. Single precision floats only have 23-bit mantissas, double precision floats have 52-bit mantissas.

Allow versioning: As with StationXML.

More data quality flags, with clear hierarchy. Data quality flags are the only clear way to distinguish between levels of data processing, but the choices are too limited. Additional data qualities that cannot currently be specified are: Data directly translated from another format , or data for which the header values have been changed, but not the data itself. A possible hierarchy would be (new in italics):

- “D” : The state of quality control of the data is Indeterminate
- “T”: Translated Raw Waveform Data from another initial format
- “R”: Raw Waveform Data with no Quality Control (reserved for SEEDlink)

· "H": Quality controlled Data, processes have been applied only to the headers
· "Q": Quality controlled Data, some processes have been applied to the data (does this mean time-series values)?
· "C": Quality controlled Data, No processes applied to time-series or header
· "M": Data center modified, time-series values have not been changed

Processing information

OBS data may go through a number of steps before being ready for archival at data centers. These processing steps should be well documented, so that any mistakes can be traced and corrected. The most obvious example is for the timing corrections, but other steps may also be useful. One possibility is to create 'opaque' miniSEED files with this information. Another would be to provide a text file (perhaps structured, such as JSON) with this information. The text or structured file would be more readable, whereas the opaque miniSEED file fits in some data structures (such as SeisComp3 data structure). An external technical report published with a persistent identifier and linked form the metadata can be the right extension to accommodate all the additional information that cannot fit in today's metadata format.

4.5 Integration in EIDA

OBS data from the AlpArray project are being archived in EIDA data centers following the guidelines described in the previous sections. Guidelines will be updated in synergy with changes to data and metadata formats as well as following new needs of the community.

Archiving data and preparing metadata

Minimum
- Waveform data archived with nomenclature proposed above as well as where possible including both time corrected and uncorrected data. Uncorrected (raw) data can be also linked via DOI metadata to the corrected data and metadata.
- StationXML created with conventions as indicated in Short Term Solutions.

Optional
- Prepare Technical Reports published with DOI and linked to the data set. These reports may allow extensive description of the data sets for advanced use cases including all the processing chain.

Data distribution

- Data discoverable and accessible using standard fdsn webservices / portals

5 Appendix C: Catalogue and Event Types in QuakeML

main author: Philipp Kästli / community expert: Andres Heinloo

5.1 Challenge 1: Event types

The current event type list

Has no defined mapping of some types as subtypes of others, and, in some cases, competing classifications (e.g.: can an "earthquake" be an "induced or triggered event"

Is not complete or not specific enough for some application cases (.e.g: volcanic tremors)

Short term solution:

Use QuakeML extension points to define your own classification in tags from your own name space.

Shortcomings:

new/alternative classifications are not part of the standard and understandable typically only for their authors; standard software will just ignore them

Long term solution:

Have types of things (such as types of events) represented in SKOS vocabularies. These allow to define class-subclass and other relationships between terms. These relationships are machine-readable and allow software to easily extract e.g. "(all types of anthropogenic) explosions, even if some events are tagged as explosions, but others more specifically as quarry blasts.

Have (in addition) event type being of type "classification" whereas a classification comes with attributes "concept" (string, the actual class i.e. SKOS term), conceptSchema (resourceIdentifier, the ID of the classification, and, potentially, classificationSource (Literature Source of the classification). Besides the ones defined with QuakeML, users could (if required) use their own conceptSchema without breaking the standard.

Have (in addition) the event type in n multiplicity, allowing an event to be classified following multiple type classifications at the same time.

The long term solution requires a new version of QuakeML Basic Event Description, and, for operative usage, of the FDSN event web service standard (-> type selection parameters, response format)

5.2 Challenge 2: Event catalogues

While QuakeML always had the intention to describe seismic event sets (or catalogues), the set was not represented by an entity, but just by the context (e.g., in XML: the file is the catalogue). This has multiple shortcomings:

entities used in different catalogues (e.g. picks, but also origins and events) need to be duplicated.

If the data model is used for other representations than xml (e.g.: databases), requiring multiple of those is not an adequate technical solution if managing mutliple, potentially overlapping catalogues.

Relationships between different catalogues are not easy to find and maintain.

There is no option to have catalog meta-information (creationinfo, name, etc.).

There is no obvious implementation for the catalogue selection request parameter in FDSN station web service.

Short term solution:

Defining catalogues implicitly by properties of e.g. events (e.g.: all events with creationinfo.agencyid=SED define the SED catalog) is possible in some cases. However, this is not transparent to the user, and solves only point 2 and 3 (5 with specific software implementations) of the issues above.

6. Appendix D: Volcano and Near Fault Observatories geophysical datasets

main author: Jean-Marie Saurel

6.1 Data Summary and Target Community

Volcano and Near Fault Observatories both typically comprise of a wide variety of geophysical measurements that monitor their objects of interest: volcanoes and active faults. Those observatories are usually set-up for very long term observations measurements. One of the main datasets produced by both observatories are classic ground motion measurements using seismometers.

A main scientific objective of those data is to understand the various processes that happen around those active objects. In order to have a complete view, a large variety of geophysical sensors are deployed on and around volcanoes and faults. Volcano observatories also play a key role in monitoring and trying to provide early warning of eruptions.

Given the wide range of data acquired, there is a wide range of users, mainly scientists spanning many different fields. Generally speaking, members of the community have their own methods for data organisation in either databases or directory structures and many don't have any metadata available. Hence the question of metadata standard for volcanological data is a critical issue to tackle. Most of the users usually deal only with basic ASCII data representation and have few or no standard to describe metadata.

Data are often continuous, evenly sampled time-series from a permanent site. Thus, generally speaking, sensors are plugged to a field digitizer to produce time-stamped data recorded in different formats. Thus, a pragmatic approach consist in using the well-established standards of the seismological community to provide data and metadata standard for volcanological and near fault observatories geophysical data.

Moreover, tools should be provided to facilitate the data access to those users, such as webservices providing instrument correction and converting data in ASCII files, because users are not yet familiar with SEED standard.

Finally, volcano and near fault observatories communities must deal with data sampling are always higher rate. A sample per minute or every 10 minutes was the standard 10 years ago although geophysical data are now sampled every second or faster (10Hz). Thus volcano and near fault observatories communities have to consider caveat on management of bigger quantity of data.

6.2.Current challenges / shortcomings

Most of the dataset is already described in the SEED manual.
However, some challenges still exist :
1. Lack of agreed guidelines for data archival and metadata formats
-> propose to use existing seismological standards for short term and long term solution
-> suggest better definition of metadata, specifically in SEED Manual Appendix A (some channels code are poorly defined)
2. Lack of easy to use metadata editor tool for seismological formats
-> propose to begin using existing tools (PDCC toolkit and dataless to stationXML converter) for short term
-> long term solution is already underway

3. Issues for polynomial sensor responses
-> stationXML solutions already approved by FDSN, implementation proposed since 2015
-> long term solution to plot polynomial response and correct data from this response type is needed (e.g. something similar to evalresp and plotresp for frequency responses)
4. End-user specific dissemination tools are missing
-> long term solution required

We recommend that the volcano and near fault monitoring communities use existing seismological standards for data storage, metadata management and data distribution for appropriate datasets. The data should be labeled according to FDSN standards (Seed naming convention, stationXML metadata).
In this case, data and metadata will be compatible with EIDA infrastructure and could be distributed either by existing or new dedicated EIDA node.

6.3.Short term solutions

- Solutions within confines of existing data format

6.3.1 Metadata Modifications

We use as basic starting point the existing Seed standard for channel naming, and stationXML for metadata information.

It is important to quantify the relationship between the input ground motion and the output digital values. In the most simple case, this is a simple sensitivity value. In other cases, common in sensors in this community, the response is described with a polynomial equation directly linking physical input value to digital output values. In most cases, a simple first order polynomial equation is sufficient (Y=a*X+b). StationXML 1.0 states it is mandatory to have a StageGain for every ResponseStageType, including the Polynomial one. This has no physical sense for a polynomial response.
A Pull Request that solves this error exists since 2015 on the FDSN github stationXML.xsd repository.

For both channel naming conventions and metadata, we provide suggestions for best practice that can be followed by the EPOS volcano and near fault communities.
For the polynomial gain stage, we don't recommend adding a sensitivity stage gain for any polynomial stage and response while waiting for the solution to be finally adopted by FDSN.

We propose to add the orientation code 'M' for magnetic field sensors type 'F' to correctly identify the Modulus of the magnetic field vector.

6.3.2 Conventions within Existing StationXML

Where relevant we follow the Seed naming convention – see Appendix A of https://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf
The following highlights key fields in StationXML that should be followed for each monitored structure.

Network.Code (2 Characters) : use the same network code as the one used for the seismic stations operated on the volcano or the near fault observatory. For example, use PF for any sensor on Piton de la Fournaise volcano, use CL for any sensor on Corynth Rift Laboratory. Must be registered with FDSN http://www.fdsn.org/networks/

Station.Code: (2-5 characters) name of the the station. Standard practices are to use a single station name for any instruments that are within a 1km diameter. If a seismic station already exist within that

range, its station code should be re-used. The station code should carry only geographic information and no informations on the type of sensors or measurements.
ISC rules for seismic station registration should be followed.
http://www.isc.ac.uk/registries/registration/#rules

Channel.LocationCode (2 characters): the location code is used to differentiate identical sensors in the same stations (e.g. several thermal sensors for thermal flux measurement). The location code can be opaque (is not intended to be interpreted), but can also provide some information about the sensors types or positions. Number are preferred.

Channel.Code (3 characters): the SEED manual Appendix A lists instruments codes that can be used (central letter). The first letter mostly reflects the sampling rate. The third letter, reflects measurement orientation or measurement details, depending on the instrument. Here follow a list of second and third letter recommendations.
<u>Tiltmeter</u> : A[E,N,0,1]
E or N if axes is within 5° from true geographic East or North, 0 and 1 otherwise.
Azimuth is along tilt direction, dip is 0.
<u>Creepmeter</u> : B[O,C,R,L,U,D]
The orientation follows Aki and Richards (1980) conventions. The strike (horizontal angle) and the slip (vertical angle 0-90°) define the orientation of the fault with regards to geographical North. The azimuth must be chosen so that the dip of the fault plunges on the right-hand side. For example, if the fault is oriented on a line at 15° from North, and the dip is on the left, we change the azimuth to be 195° from North and the dip comes to the right hand. Most of the time, the dip at the surface is unknown and in this case considered as 90°.

- ?BO : fissure Opening
  - Azimuth : fault direction + 90 (perpendicular to fault)
  - Dip : 90 if unknown, actual value otherwise
  - Increase in data is a fault opening
- ?BC: fissure Closing
  - Azimuth : fault direction + 90 (perpendicular to fault)
  - Dip : 90 if unknown, actual value otherwise
  - Increase in data is a fault closing
- ?BR: Right for dextral (right lateral) movement
  - Azimuth : fault direction
  - Dip : 90 if unknown, actual value otherwise
  - Increase in data is a dextral motion
- ?BL: Left pour senestral (left lateral) movement
  - Azimuth: fault direction
  - Dip : 90 if unknown, actual value otherwise
  - Increase in data is a senestral motion
- ?BU: Upward movement
  - Azimuth: fault direction
  - Dip : 90 if unknown, actual value otherwise
  - Increase of data is an upward motion of the right-hand side of the fault
- ?BD: Downward movement
  - Azimuth: fault direction
  - Dip : 90 if unknown, actual value otherwise
  - Increase of data is a downward motion of the right-hand side of the fault

<u>Pressure</u> : D[O,I,D,F,H,U]

- ?DO : air pressure outside (outside building, outside container)
- ?DI : air pressure inside (inside building, inside container)

?DD : pressure down hole, at the bottom of a well (water pressure)
?DF : dynamic air pressure (infrasound)
?DH : water pressure (hydrophone, underwater pressure sensor)
?DU : underground pressure (pore pressure at some depth)

Magnetic field : F[Z,N,E,M]
?FZ : vertical component of magnetic field
Azimuth : 0
Dip : 90 (positive for downward magnetic vector)
?FN : North component of magnetic field
Azimuth : geographical azimuth of the local magnetic North
Dip : 0
?FE : East component of magnetic field
Azimuth : geographical azimuth of the local magnetic East
Dip : 0
?FM : Modulus, scalar value of the magnetic field vector

Humidity : I[O,I,D,?]
?IO : outside environment humidity
?II : internal humidity
?ID : Down Hole humidity
?I? : humidity inside cabinet or any other mnemonic letter

Temperature : K[O,I,D,?]
?IO : outside environment temperature
?II : internal temperature
?ID : Down Hole temperature
?I? : temperature inside cabinet or any other mnemonic letter

Water Current or Flow Rate : O[O,D]
?OO : outside water flow rate (thermal spring, river)
?OD : Down Hole water flow rate
Azimuth : geographical azimuth of the main flow direction

Electric Potential : QU

Rainfall : RO
Outside Rainfall, no azimuth and dip.

Tide : TZ
Vertical (always) local free water high.

Cloud cover : UO
Outside cloud cover, can be used to assess volcanic ash clouds.

Volumetric Strain : V[A,B,C]
Azimuth : according to instrument and axes orientation, relative to geographical North
Dip : 0

Wind : W[S,D]
?WS : wind speed
?WD : wind direction
No dip and azimuth

6.4.Long term solutions

There is very little to do in order to integrate those geophysical data into existing seismological community format. What could be done are tools to facilitate the production and use of stationXML metadata for volcano and NFO communities that are not very used with this format.

StationXML editor

An easy to use stationXML creator and editor tool with a user friendly GUI is currently under development and is expected for release within a couple of years.

Polynomial response plot and correction

Most, if not all, of the geophysical data measurements from Volcano and Near Fault Observatories are described with a polynomial stage at some point in the response. Even if this response stage is supported in the stationXML metadata standard, none of the current tools (ObsPy, SAC, evalresp) can use the informations to correct the raw data and produce physical data. The same happens when we want to plot an actual polynomial response to check if it's correct.

This greatly prevent the use of the data as users are then required to dig inside the metadata and correct themselves with mathematical operations the data.

Specific dissemination tool

The majority of Volcano and Near Fault Observatories geophysical data user have no experience to deal with existing seismology community tools. Thus data should also be available in various common formats, such as ASCII, CSV or Matlab files and with the instrument response corrected.

This could probably be achieved with the new FDSN timeseries webservice with some minor additions and a nice an easy to use Web interface/URL builder.

6.5 Integration in EIDA

Since all those geophysical data already meet existing FDSN standards, there would be compatible for integration within EIDA, either via existing nodes or via a new dedicated node.

7 Appendix E: Volcano and Near Fault Observatories geochemical datasets

main author: Jean-Marie Saurel

7.1 Data Summary and Target Community

Volcano and Near Fault Observatories acquire a wide variety of geochemical measurements to monitor volcano activity and fault activity respectively. Those observatories are usually set-up for very long term observation. For the last 30 years, gases and water were sampled manually at a more or less regular frequency and analyzed afterward in the laboratory. Since a few years, technology has evolved and some gas sensors have became readily available with sufficient accuracy and reliability to allow automatic, regular sampling.

One of main scientific objectives of those data is to understand the various processes that happens inside the volcano and chemical analysis of gases and waters offers a very interesting insight on the underground lava related processes.

Volcano observatory play also a key role in monitoring and trying to provide early warning of eruptions.

Geochemists usually deal with tabular data or ASCII files of historically very little volume.

7.2.Current challenges / shortcomings

Currently, there are no widespread community agreed data and metadata format for geochemical data exchange. Tabular data are most of the time self explanatory and methods and how the measure were obtained are in referenced scientific papers.
For the emerging continuous automatic gas sampling, there are no standard data or metadata formats. SEED standards could be used as the gas concentration / flux data being collected is evenly sampled, and is more or less continuous.

However, the current SEED standards does not address the geochemical community needs :
1. Lack of appropriate stream definitions
The current SCNL stream definition doesn't allow to define a flux or concentration measurement of given gas species. For ease of use within the community, the channel code should be explanatory and should contain the chemical species code. Note the seismological community has become accustomed to using relatively cryptic channel codes, so this can be overcome.
-> long term solution could come from Next Generation Miniseed (also called miniseed3)

2. Lack of solution to address uniformly both manual and automatic sampling
Some chemical species concentration or flux have been monitored manually for decades and are now beginning to be monitored automatically. It is very important that end-users can manipulate easily long term time series of chemical measurements. A common solution for both uneven manual analysis and automatic analysis would be welcome.
-> long term solution still to be found : SEED standards are currently not made for uneven data sampling

7.3.Short term solutions

7.3.1 Metadata Modifications

Not applicable.

7.3.2 Conventions

Not applicable.

7.4.Long term solutions

Next Generation Miniseed could possibly meet the requirements of automatic continuous gases and water chemical analysis.
However, manual analysis (historic and contemporary) would not be able to fit in this format, still targeted toward evenly sampled data.

Furthermore, as technologies evolve, it’s almost impossible to foresee what chemical species are going to be done in the future on any given volcano. The identification of the streams and data must then be kept open to addition of any new chemical specie or analysis type (flux, concentration, isotopic content ….). A future version of miniSEED should include the option to have a 'living' 'Appendix A' where with relatively little overhead, new data types can be added

7.5 Integration in EIDA

Not applicable at the moment.

8 Appendix F: Moving Sensors

main author: Helle Pederson

8.1 Data Summary and Target Community

Moving sensors are of relevance in several communities that use FDSN standards for data distribution, and data distribution is effectively increasing. This is particularly the case for permanent or semi-permanent observations on landslides, for installations on ice (whether on glaciers or slowly moving ice sheets), at sea (in particular for free-floating instruments) and to a lesser extent on volcanoes. Increasingly, data from temporary experiments from such sites are also distributed.

8.2.Current challenges / shortcomings

Presently the FDSN standards cannot appropriately handle moving sensors. There is no way to document the motion, and there is even no indication the sensor is not in a fixed position over time. This practically may have a large scientific impact in terms of data analysis that use relative position of sensors, or on the scientific interpretation of results, as the local parameters may not be known at the new position. Finally, some key derived products from the raw waveform data, such as earthquake locations, depend on the sensor coordinates being correct at a particular time.

8.3.Short term solutions

- The position of the sensor (in absolute or relative terms) could be included into the comments field of stationXML. The upside is the simplicity of implementation. The downside is a cumbersome metadata file and potentially underused information.
- Several location codes could be used, and the scientific user can then interpolate between position data points. This ad hoc solution could be considered in very simple cases, but to make a new use of the location code field will have to be properly documented to promote best practices. Information of position between each new location code would be lost.

8.4.Long term solutions

- Allow for an additional channel specification with continuous (but not necessarily evenly sampled?) information on position within the data itself, and not only in the metadata
- If the equipment is associated with a good GPS antenna equipment, distribute the position independently via the proper datacenters, tools and formats. The challenge of connecting the two data streams from different data centers might be a severe practical limitation of this solution.

8.5 Integration in EIDA

Solutions that are based on present standards can be integrated into EIDA without any impediment. Practically datacenters based on SeisComp3 will need to cooperate to check that use of location codes or comments as suggested are properly taken into account.

9 Appendix G: Strong Motion Records without Precise Timing

main author: Christos Evangelidis

9.1 Data Summary and Target Community

Global and national strong motion databases contain earthquake waveform records for dissemination to the geotechnical and structural engineering communities (e.g. PEER, COSMOS, ITACA etc).

Despite the often poor quality of the recording devices, incomplete metadata and often missing timing, these datasets include near-source recordings from significant earthquakes, and are thus some of the most valuable in the seismological community, and are crucial for the strong motion and engineering community. Data is often old, so existing SEED naming standards widespread today were only partly used or not used at all. It is important that these datasets are integrated into modern archives as far as possible using sensible conventions that do not lose the provenance of the data.

Nowadays, the accurate timing of seismic and strong motion data (both continuous and triggered events) is secured with millisecond precision using GNSS timing sources (GPS, GLONAS, BeiDou) or network time synchronization (PTP, NTP). This absolute timing is in contrast to the non-precise (if any) timing of older generation strong motion instruments. These instruments have a timing that is updated and synchronized during maintenance visits. Records for such instruments are typically distributed with a timing relative to the event origin time.

The need of the seismological community to integrate such records alongside the continuous broadband and strong motion records is clear, so it is sensible to include these datasets into the existing archives and use the same data / metadata formats. Conventions are required though to alert users to and/or to account for the imprecise or absent timing.

Older triggered strong motion records can be added to EIDA as day-long mseed files with larger than normal gaps. This is demonstrated by the Swiss triggered strong motion records from the 1990's that are included in EIDA following this procedure.

Tools should be provided to facilitate the data access, such as webservices providing instrument correction and converting data in ASCII files and vice versa.

The ESM (esm.orfeus.eu) already includes and provides access to many of these datasets.

9.2.Current challenges / shortcomings

1. Lack of agreed guidelines for data archival and metadata formats
   -> propose to use existing seismological standards for short term and long term solution
2. When seismological standards are used, there is no standard convention / best practice
   -> short term solution addresses this
3. Additional and important timing specific information is missing (eg record timing absolute or relative)
   -> long term solution proposes an extension to stationXML
4. End-user specific dissemination tools are missing
   -> long term solution required
5. Specific waveform conversions from various text formats to mseed format are missing

-> long term solution required

We recommend existing seismological standards for data storage, metadata management and data distribution, with appropriate extensions. The data should be labeled according to FDSN standards (Seed naming convention, stationXML metadata) and should be archived and distributed using the EIDA infrastructure. However, strong motion records with imprecise timing instrumentation cannot be sufficiently described using the existing stationXML. Additionally, though each agency has its own naming convention, there is not a consistent application of the standard by managers maintaining these data.

9.3.Short term solutions

9.3.1 Metadata Modifications

We use as basic starting point the existing Seed standard for channel naming, and stationXML for metadata information. For both channel naming conventions and metadata, we provide suggestions for best practice that can be followed by the EPOS community.

9.3.2 Conventions

Where relevant we follow the Seed naming convention – see Appendix A of https://www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf
The following highlights key fields in StationXML that should be followed for each monitored structure.

Network.Code (2 Characters) : use the network code of the seismic network that installs and manages the data, eg HL for Hellenic Seismic Network. Must be registered with FDSN http://www.fdsn.org/networks/
If no network code exists for a defunct network, adopt an appropriate registered network code from an affiliated group, eg university or national / regional network

Station Information
Station.Code: (2-5 characters) name of the station or structural array being monitored. Station.code name should not necessarily be registered at ISC, since it is quite possible the name is used by another network, and changing the most basic identifier of a significant record can lead to confusion.

Location Information
Option A1:
Channel.LocationCode (2 characters): In case of stand-alone stations without pre-existing location codes, the location code becomes

-TR (T: Trigger, R: Relative) for triggered stations without any independent time stamp
-TA (T: Trigger, A: Absolute) for triggered stations with accurate timing

Option B:
Channel.LocationCode (2 characters): In case of structural monitoring array the approach proposed in Appendix B is followed

Channel Information:
Channel.Code (3 characters): for strong motion sensors, we assume we are collecting (triaxial) accelerometric data, so channel name should be

HN[Z23] for orthogonal sensors with horizontal sensors oriented in the principal directions of the structure (standard practice in structural monitoring),
HN[ZNE] for orthogonal sensors oriented in traditional components with respect to North.
The azimuth of the sensor to North is recorded in the stationXML. Note the azimuth can change from sensor to sensor, and the individual dip and azimuth for each sensor are defined in the metadata.
Channel.Type: “TRIGGERED”
Channel.ClockDrift: “9999999999”
Comment.Value: “Timing relative to event”
Comment.Author: Possibly the fdsnws_event service that is attached to the record

The MiniSEED data record itself should also have a data quality flag indicating that the time tag is questionable and an I/O-clock flag indicating unlocked clock.

9.4.Long term solutions

Modifications to StationXML: Option in Station XML to provide a timing flag (A: Absolute, R: Relative). Must be at epoch level as GPS timing can be added to a station later.

9.5 Integration in EIDA

I. Archival and metadata
Waveform data - archived with SNCL as proposed above
StationXML - created with conventions as indicated in Short Term Solutions.

II. Distribution
Data discoverable and accessible using standard FDSN webservices / EIDA portal. ESM for European datasets.

10 Appendix H: New Communities and Related Challenges

main authors: Helle Pedersen, Angelo Strollo, John Clinton

10.1 Data Summary and Target Community

The last decades has been characterised by large amount of new additional permanent and temporary conventional seismic stations becoming available to the seismological community. It is already clear there are some game-changing technologies under development in the seismological community. The demand for new dense observations using new technologies is advancing fast in seismology. This includes Large N deployments consisting of huge numbers of easy-to-install geophones; cheap sensors (e.g. MEMs accelerometers) used by mobile devices as well as for infrastructure monitoring; and fibre-optic based technologies, that are each starting to show their great potential in providing quality data with a wide spectrum of applications ranging from Tsunami early warning to Infrastructure monitoring. Although data quality and resolution of the three groups mentioned above are different, for seismic monitoring and archival infrastructures, these groups have in common the potential to produce large volumes of data in a very short period of time due to both the extremely dense spatial and temporal resolutions. This appendix will only list briefly the current status and possible strategies to handle these data. Additional details on the future long term solutions are expected in SERA Deliverable 4.7 (Strategies for future network design) that will be produced within task 4.4 at M36 after a dedicated discussion with the stakeholders.

- Large N deployments are becoming popular thanks to the availability of cost/power effective instruments. Typically of short duration, they may imply a very varying number of sensors (from a few tens to thousands), and are presently carried out with high frequency sensors. Some manufacturers have only high frequency sampling (for example, a minimum sampling rate of 250Hz). Field installations are easy, and are often carried out by teams of non-specialist personnel.
- Cheap sensors are already used since about a decade in educational projects showing capabilities to provide satisfactory recordings for earthquakes. Recently a number of crowdsourcing projects have been launched making available a growing number of additional sensors from private users to the seismological community. These crowd-sourced projects include tailor-made cheap sensors and digitizers for private consumers as well as usage of existing sensors in mobile devices. In both cases the easiness to access new data, sometimes in area where conventional seismic stations are not available, is making these data valuable for real-time applications in the context of rapid response and earthquake early warning.
- Fibre-optic can be used as seismic arrays measuring ground motion from earthquakes for example using the so called distributed acoustic sensing technique (DAS) with existing underground fibre-optic cables (normally used for telecommunications, such as internet, television, and telephone service) up to tens of kilometres long or deploying dedicated short fibre-optic cables. Alternative usage of fibre-optic communication cables can be also to place along their repeaters (typically 50-100 km intervals) conventional sensors and only use the fibre-optic for communication for example at the ocean floor using existing submarine cables. The former allows for example to image the internal structure of faults with high resolution as well as to infer creeping processes of faults at sub-micrometre step ((https://www.nature.com/articles/s41467-018-04860-y). The latter has the potential to provide an unparalleled global network of real-time data for ocean climate and sea level monitoring and disaster mitigation from earthquake and tsunami hazards (https://eos.org/meeting-reports/submarine-cable-systems-for-future-societal-needs)

10.2.Current challenges / shortcomings

Besides the issues related to the large volume of data produced by these new technologies (a DAS acquisition can produce more than 50 TB of data for ~1 month at the 1-2 km scale) the main challenge at the moment is to handle the large number of sensors (virtual in the case of DAS) within the current standard FDSN station naming conventions that affects both data and metadata formats. The discussion on possible exertions to the naming conventions is currently in progress within the FDSN with proposal for new or extended data and metadata formats

10.3.Short term solutions

Several Large-N experiments have been already carried out in seismology and archived at some FDSN data centers using the actual standards. Data from these experiments are currently distributed via standard FDSN services but with clear limitations due to the fast growing number of experiments as well as additional DAS data expected to be produced in the near future.

10.4.Long term solutions

- Review the current metadata and data formats at the FDSN level (currently in progress in WG III)
- Consider alternative formats for large volumes of data suitable also for ingestion at HPC facilities e.g. PH5 (HDF5 format) possibly adding a new output format to the current FDSN services while preserving the miniSEED format output.
- Consider a revision of the metadata to accommodate non-conventional seismic sensors and in particular to the DAS applications consider the description of virtual sensors along the fibre-optic cables.

10.5 Integration in EIDA

With limited capability due to the current standards Large-N experiments and cheap sensors (educational networks) are being integrated already in some EIDA data centers. Challenges related to the naming conventions description of the sensors and to the data volume need to be addressed to provide a satisfactory user experience.