

A Strawman Proposal for OBS data/metadata preparation

Motivation

- OBS data are different from land data, need to account for these differences while making them as easy as possible to use by general public
 - Add rules for data and metadata
 - Provide tools for OBS-specific processing
 - Sensor orientation
 - Clock validation/recorrection
 - Noise removal
 - Extraction of data in active seismic format
 - Clearly document these rules and tools

Document

- First version made available in DOCX format
- Structure
 - Rules for data files (miniSEED)
 - Rules for metadata files (StationXML?)
 - Guidelines for pre-processing tools
 - User guide

Rules for data files

- Assume miniSEED (only format I know that allows fine-scale time corrections)
- Timing corrections (because clock drifts at seafloor)
 - Recommendation
 - Apply the time correction to each miniSEED record header (generally every 4096 bytes, so every 40 seconds for efficiently compressed 24-bit data at 100 sps), specify the correction applied and the fact that it was applied in each header. Also, set the “data quality” field to “Q”. If there is no information about clock drift but there is a risk of clock drift, set the “data quality field to “M”.
 - Fields concerned:

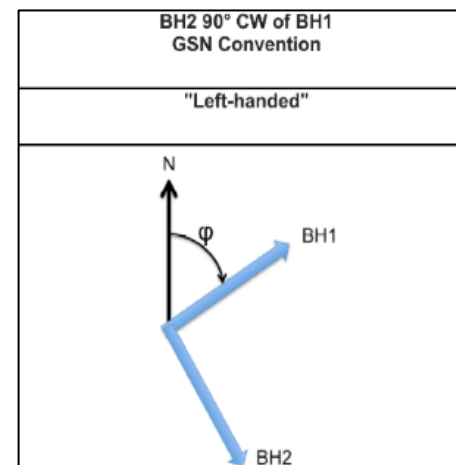
Field	Bytes	Name	Comments
2	7	Data header/quality	“Q” for clock drift corrected, “M” for possible clock drift not corrected
8	21-30	Record start time	Clock drift corrected time
12	37	Activity flags	Set bit 1 to 1
16	41-44	Time correction	.00001 seconds that were ADDED to the initial time to obtain the corrected time

Rules for data files

- Timing corrections (because clock drifts at seafloor)
 - Recommendation
 - The 'qedit' software can apply a linear clock correction to miniSEED data, taking care of Fields 8, 12 and 16. I have heard some concerns that 'qedit' is not a standard miniSEED tool but I have not had any problems and I am not aware of standard software that can do a linear clock correction. There may be others that I am not aware of. The msmod software can be used to set the data header/quality field.
 - Make sure standard softwares (for stitching data sets together, or creating spectrograms) take into account individual record headers
 - Alternatives:
 - Put the time correction in the header but do NOT apply it
 - May help with stitching data sets together
 - May hurt pick times
 - Specify sampling rate exactly (accounting for drift)
 - Precision (8 significant digits) not possible with current data formats
 - Resample data at specified sample rate
 - Waveform distortion problems?

Rules for data files

- Pressure channel names
 - Currently (IRIS) names all pressure channels « ?DH », which means « hydrophone » according to the SEED manual
 - Call differential pressure gauge channels « ?DF » (« infrasound ») and absolute pressure gauge channels « ?TZ » (« tide gauge »)?
 - Or add other pressure channel “orientation codes” to SEED standard?
- Channel orientation codes
 - Use “1” and “2” if the instrument is leveled and orthogonal but the “NS” and “EW” axes are not along geographic coordinates
 - Use GSN standard (“left handed”)
 - “N”->1
 - “E”->2



Rules for data files

- Site names for repeat deployments
 - If instruments are redeployed at same spot (but from the sea surface, so some drift), can we give them the same station name (with updated seafloor coordinates)?
 - IRIS does not repeat station names: they increment the last alphanumeric character at the end of the station name for repeated deployments
- Information on processing steps
 - In opaque miniSEED files?
 - Fit into some archive structures (i.e., SDS), but may be rejected by data centers
 - As text files?
 - Is there a standard place to put this (with state of health? Preamble directory? Catch-all directory?)

Rules for metadata files

- Do we need to bother with dataless SEED? Or go straight to StationXML?
- OBS-specific information
 - Water level
 - useful for removing/exploiting water surface reflections.
 - Generally 0 (sea level, default value?), but not if deployments are made in lakes.
 - “water level” seems simpler to implement than “water depth” because the default value would be “0” rather than “-elevation”.
 - Estimated position errors (provided for in StationXML)
 - Comment on the method used to position the OBS
 - Surface release position
 - Near-bottom release position
 - Airgun survey
 - Acoustic survey
 - If horizontal channels are not geographically oriented, their “Azimuth” field should be set to “0”

Software

- Should all work seamlessly on data-center supplied data
- Clock drift confirmation
 - Based on hypocenter time residuals
 - Based on noise correlation
- Noise removal
 - On vertical channel, caused by compliance and tilt
- Sensor orientation
 - Several tools available based on EQ back-azimuths
 - One tool already online (<http://www.obsip.org/data/obs-horizontal-orientation/>)
- Common pipeline for data preparation
- Active seismic data extration

User manual

- OBS-specific information
 - Information about
 - clock drift and how to identify if your data is drift corrected
 - orientation standards and determination
 - Noise removal
 - Data redaction by military
 - Pressure channel uses
 - Pressure channel names
 - Processing codes