OBS services in EPOS

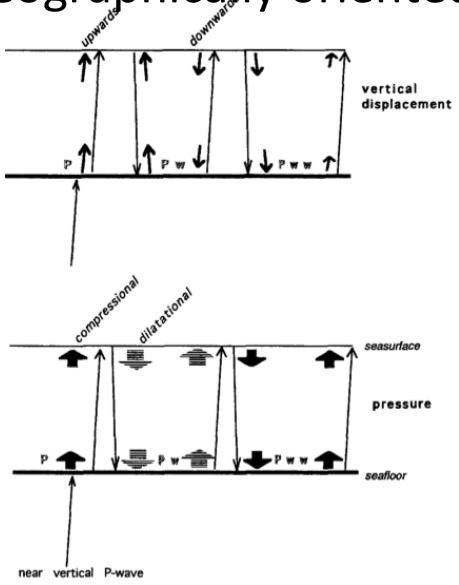
Wayne Crawford

No web services (yet)

- Our goal is to make OBS data as compatible as possible with "standard" seismological data
  - Use their web services
- OBS-specific web services?
  - Depends on OBS-specific properties
  - We are working on contours of these properties and ways to communicate/store them

What is different about OBSs?

- Recording
  - Non-standard loggers (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 deployment
- Sensors
  - Horizontal channels generally not geographically oriented
  - Pressure sensors
- Noise
  - Seafloor currents
  - Motion under ocean waves
  - Strong sea-surface reflections



Blackman et al., 1995, BSSA

Challenges in distributing OBS data

- 1: Creating the 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
    - NRL libraries (based on RESP files) do not completely inform StationXML
  - Mobile instrument problem:
    - 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)

Challenges in distributing OBS data

- 2: Making the data as useful as possible
  - Reorienting the horizontal sensors
  - Verifying clock corrections
  - Removing current and ocean wave noise
  - Teaching seismologists about using pressure sensors.
    - Remove noise
    - Remove surface reflections
    - Studying crustal structure with seafloor-specific techniques

Creating the data and metadata

- We are developping a system to generate clock-corrected data from basic miniSEED data
  - Facilities must convert their data to miniSEED, but do not need to have final network/station names, nor correct the clock drift
  - Eases the load on parks, assures the clock drift is documented and always treated the same way.

- We are developing a three-part information system for informing the data and metadata creation:
  - 1: **Instrument information** (provided by OBS facilities)
  - 2: **Network/stations information** (provided by OBS facilities after campaign)
  - 3: **Campaign information** (provided by the reference scientist)

Creating the data and metadata

- **Future**: make web-service and graphical interface
  - Could enter information graphically or through formatted information files
  - Database behind is invisible to users, but would allow previously-declared instruments/components to be selected
- **Present**: information files using a standard machine- and human-readable format
  - Should be easily readable into common programming data structures (lists and associative arrays [also known as « maps » or « dictionaries »])
  - Currently using YAML
    - Easy to read
    - Efficient specification of values common to all instruments

Campaign information file

- One file per campaign
- Filled in by reference scientist
- Indicates
  - Reference scientist information
  - OBS facilities used
  - network information
  - Verification tool information (currently time windows in which to plot waveforms)
  - Ancillary information
    - Expeditions
    - ….
- Is not needed to create data or metadata, but allows verification of information provided by the other files, as well as a means of providing feedback/verification to the reference scientist

```
format_version: “0.9”
campaign:
    information_version: “2017-10-31 WCC”
    reference_name: “EMSO-MOMAR2016”
    reference_scientist:
        name: ”Albert Einstein"
        institution: "Institut de Physique du Globe de Paris"
        email: “einstein@ipgp.fr”
        telephone: ~
    OBS_providers:
        "INSU-IPGP OBS park":
            email: parc-obs@insu.fr
            representative: "Wayne Crawford <crawford@ipgp.fr>”
    network:
        code: 4G
        start_date: 2007-07-01
        end_date: 2025-12-31
        FDSN_registered: True
    verification:
        waveform_samples:
            -
                date: ~
                duration: ~
                title: ~
            -
                ....
    ancillary:
       expeditions:
            -
                name: ~
                ship: “Nina”
                start_date": ~
                end_date": ~
                comments: "Deployment"
            -
                name: ~
                ship: “Pinto”
                ....
```

Instrumentation information file

- One YAML file plus directory of response files per OBS facility
- Provides StationXML compatible equipment/ response information
- Can be combined with network.yaml file to create StationXML
- Largest (but least modified) of the information files
- Can also be used to generate RESP files for NRL

1: Global information and variables

```yaml
instrumentation:
    information_version: "2017-09-15 WCC"
    facility:
        reference_name: "INSU-IPGP"
        full_name: "INSU-IPGP OBS Park"
        email: "obs_ipgp@insu.cnrs.fr"
        website: "http://parc-obs.insu.cnrs.fr"
        director: "Wayne Crawford"
        chief_engineer: "Romuald Daniel"
        phone_number: ""
    response_format: "DBIRD"
    response_directory: "DBIRD_OBS_INSU-IPGP"
    variables:
        # VARIABLES TO BE SPECIFIED IN THE NETWORK FILE
        # AND THEIR DEFAULT VALUES
        sample_rate: "62.5"
        serial_number: ~
        digitizer_serial_number: "generic"
        analog_filter_serial_number: ~
        pressure_serial_number: ~
        pressure_manufacturer: ~
        dpg_calibration_code: "generic"
        seismometer_serial_number: ~
        seismometer_calibration_code: "1-399"
        # THESE ARE FOR MULTIPLE HYDROPHONES ON A STATION (HYDROCTOPUS)
        pressure_serial_number_1: ~
        pressure_serial_number_2: ~
        pressure_serial_number_3: ~
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- Can also be used to generate RESP files for NRL

2: Building blocks: dataloggers

```yaml
dataloggers:
    LC2000:
        digitizer:
            type: "delta-sigma A/D converter"
            description: "CS5321 delta-sigma A/D converter"
            manufacturer: "Cirrus Logic"
            vendor: ~
            model: "CS5321"
            DBIRD_file: "digitizer/Scripps#LCPO2000_CS5321##theoretical#"
            serial_number: "{digitizer_serial_number}"
        digital_filter:
            type: "FIR digital filter chip"
            description: "CS5322 digital FIR filter"
            manufacturer: "Cirrus Logic"
            vendor: ~
            model: "CS5322"
            DBIRD_file:
                "dig_filter/Scripps#LCPO2000_CS5322#{sample_rate}sps#theoretical#"
            serial_number: "{digitizer_serial_number}"
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- Can also be used to generate RESP files for NRL

3: Building blocks: analong filters

```yaml
analog_filters:
    DPG_CARD:
        type: "DPG_Card"
        description: "Differential Pressure Gauge Card"
        manufacturer: "SIO-LDEO"
        vendor: ~
        model: ~
        DBIRD_file: "ana_filter/SIO_LDEO#DPG_Card##theoretical#"
        serial_number: "{analog_filter_serial_number}"
    BBOBS_CARD_0P225X:
        type: "Analog gain card"
        description: "INSU BBOBS gain card :0.225x"
        manufacturer: "SIO or IPGP"
        vendor: ~
        model: ~
        serial_number: "{analog_filter_serial_number}"
        DBIRD_file: "ana_filter/INSU#BBOBS#gain0.225#theoretical#"
    BBOBS_CARD_1X:
        type: "Analog gain card"
        description: "INSU BBOBS gain card : 1x"
        manufacturer: "SIO or IPGP"
        vendor: ~
        model: ~
        serial_number: "{analog_filter_serial_number}"
        DBIRD_file: "ana_filter/INSU#BBOBS#gain1.0#theoretical#"
    HYDRO_GAIN_16X:
        type: "Analog gain/filter card"
        description: "SIO gain/filter card, hydro channel (16x)"
        manufacturer: "SIO or IPGP"
        vendor: ~
        model: ~
        serial_number: "{analog_filter_serial_number}"
        DBIRD_file:
            "ana_filter/Scripps#SPOBS#HydroL22x16#theoretical#"
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- Can also be used to generate RESP files for NRL

4: Building blocks: sensors

```
  sensors:
        VELOCITY_TRILLIUM_T240_SS:
            type: "Broadband seismometer"
            description:
                 "Trillium T240 single-sided, serial number
                  {seismometer_calibration_code}"
            manufacturer: "Nanometrics, Inc"
            vendor: "Nanometrics, Inc"
            model: "Trillium T240"
            serial_number: "{seismometer_serial_number}"
            DBIRD_file:
                "sensor/Trillium#T240#SN{seismometer_calibration_code}400-
_singlesided#theoretical#"
        PRESSURE_DPG:
            type: "Differential Pressure Gauge"
            description: "Differential Pressure Gauge"
            manufacturer: "{pressure_manufacturer}"
            vendor: ~
            model: "DPG"
            serial_number: "{pressure_serial_number}"
            DBIRD_file: "sensor/SIO-LDEO#DPG#{dpg_calibration_code}
#theoretical#"
        PRESSURE_HTI_90U:
            type: "Hydrophone"
            description:
                    "HiTech HTI-90-U hydrophone with integrated preamp,
                    0.05-2500 Hz"
            manufacturer: "HiTech, inc"
            vendor: ~
            model: "HTI-90-U"
            serial_number: "{pressure_serial_number}"
            DBIRD_file: "sensor/HiTech#HTI-90U#SIO_preamp#theoretical#“
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- Can also be used to generate RESP files for NRL

3: Final product: instruments by model

```
models:
    "BBOBS1_2":
        equipment:
            type: "Broadband Ocean Bottom Seismometer"
            description:
                "LCHEAPO 2000 Broadband Ocean Bottom Seismometer,
                configuration 2: vertical channel preamp gain =
                1.0. valid from 2012-11 on"
            manufacturer: "Scripps Inst. Oceanography - INSU”
            model: "BBOBS1_2"
            serial_number: "{serial_number}"
        channels:
            "BDH:00":
                datalogger: LC2000
                ana_filter: DPG_CARD
                sensor: PRESSURE_DPG
                azi_dip: AZIDIP_DPG
            "BH1:00":
                datalogger: LC2000
                ana_filter: BBOBS_CARD_0P225X
                sensor: VELOCITY_TRILLIUM_T240_SS_A
                azi_dip: AZIDIP_SEISMOMETER_12
            "BH2:00":
                datalogger: LC2000
                ana_filter: BBOBS_CARD_0P225X
                sensor: VELOCITY_TRILLIUM_T240_SS_A
                azi_dip: AZIDIP_SEISMOMETER_12
            "BHZ:00":
                datalogger: LC2000
                ana_filter: BBOBS_CARD_1X
                sensor: VELOCITY_TRILLIUM_T240_SS_A
                azi_dip: AZIDIP_SEISMOMETER_Z
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- Can also be used to generate RESP files for NRL

3: Final product: instruments by model

```
models:
    ”SPOBS2":
        equipment:
            type: ”Short Period Ocean Bottom Seismometer"
            description:
                "LCHEAPO 2000 short period Ocean Bottom
                 Seismometer: 4 channels, L-28 3C geophone
                 and HiTech HYI-90U 30s hydrophone"
            manufacturer: "Scripps Inst. Oceanography - INSU”
            model: “SPOBS2"
            serial_number: "{serial_number}"
        channels:
            "BDH:00":
                datalogger: LC2000
                ana_filter: HYDRO_GAIN_16X
                sensor: PRESSURE_HTI-90U
                azi_dip: AZIDIP_HYDROPHONE
            ”SH1:00":
                datalogger: LC2000
                ana_filter: GEOPHONE_GAIN_128X
                sensor: VELOCITY_SERCEL_L28
                azi_dip: AZIDIP_SEISMOMETER_12
            ”SH2:00":
                datalogger: LC2000
                ana_filter: GEOPHONE_GAIN_128X
                sensor: VELOCITY_SERCEL_L28
                azi_dip: AZIDIP_SEISMOMETER_12
            ”SH3:00":
                datalogger: LC2000
                ana_filter: GEOPHONE_GAIN_128X
                sensor: VELOCITY_SERCEL_L28
                azi_dip: AZIDIP_GEOPHONE_Z
```

Instrumentation information file

- One YAML file plus response files/ directory per OBS facility
- Reproduces StationXML logger and sensor fields
- Can be combined with network.yaml file to create StationXML
- **Can also be used to generate RESP files for NRL**

4: Alternative Final product: Logger definitions for NRL

```
loggers_NRL:
    # Convenience definitions for creating Nominal Reference Library
    # loggers.  The {sample_rate} variable will affect the
    # choice of DBIRD files in the "datalogger" item
       LC2000_DPG_{sample_rate}:
           datalogger: LC2000
           ana_filter: DPG_CARD
       LC2000_HYDROPHONE_{sample_rate}:
           datalogger: LC2000
           ana_filter: HYDRO_GAIN_16X
       LC2000_GEOPHONE_{sample_rate}:
           datalogger: LC2000
           ana_filter: GEOPHONE_GAIN_128X
       LC2000_BBOBSx1_{sample_rate}:
           datalogger: LC2000
           ana_filter: BBOBS_CARD_1x
       LC2000_BBOVBSx0p225_{sample_rate}:
           datalogger: LC2000
           ana_filter: BBOBS_CARD_0P225X
```

Network information file

- One file per OBS facility and campaign
- Information with no corresponding place in StationXML are placed under the key « non-standard »
  - Includes OBS-specific information (clock synchronisation , station localization method...)

```
network:
    information_version: "0.4 (20170906_WCC)"
    network_code: "4G"
    instrumentation_file: "INSU-IPGP.instrumentation.yaml"
    stations:
        "LSVNI":
            site: "Lucky Strike volcano North"
            start_date: "2015-04-23T10:00:00Z"
            end_date: "2016-05-26T23:00:00Z"
            sample_rate: 62.5
            comment_list: []
            station_location: "00"
            instrument:
                model: "BBOBS1_1"
                serial_number: "04"
                pressure_serial_number: "IP004"
                pressure_manufacturer: "IPGP"
                seismometer_serial_number: "138"
                seismometer_calibration_code: "1-399"
            locations:
                "00":
                    latitude: 37.31960
                    longitude: -32.27909
                    elevation: -1798
                    lat_uncert_m: 20
                    lon_uncert_m: 20
                    elev_uncert_m: 20
                    depth: 0
                    geology: "unknown"
                    vault: "Sea floor"
                    non-standard:
                        localization_method: "Acoustic survey"
            non-standard:
                original_name: "I1"
                clock_correction_linear:
                    time_base: "Seascan MCXO, ~1e-9 nominal drift"
                    reference: "GPS"
                    start_sync_reference: "2015-04-22T09:21:00Z"
                    start_sync_inst: "0"
                    end_sync_reference: "2016-05-28T22:59:00.1843Z"
                    end_sync_instrument: "2016-05-28T22:59:02Z"
        "LSVNC":
                    ....
```

References to instrumentation information file

Network information file

- YAML repeated nodes streamline station descriptions

Repeated node definitions (at file top)

```
station: &DEFAULT_STATION
    sample_rate: 62.5
    comment_list: []
    station_location: “00”
```

```
Location_defaults: &LOC_DEFAULTS
        depth: 0
        geology: “unknown”
        vault: “Sea floor”

location_methods:
    loc_acoustic: &LOC_ACOUSTIC
        lat_uncert_m: 5
        lon_uncert_m: 5
        elev_uncert_m: 10
        non-standard:
            localization method: “Acoustic Survey”
```

```
linar_clock_defaults: &LINEAR_CLOCK_DEFAULTS
    time_base: "Seascan MCXO, ~1e-8 nominal drift"
    reference: "GPS"
    start_sync_instrument: "0”
```

```
network:
   information_version: "0.4 (20170906_WCC)”
   code: “4G”
   instrumentation_file: "INSU-IPGP.instrumentation.yaml”
   stations:
      ”LSVNI":
         <<: *DEFAULT_STATION
         site: “Lucky Strike volcano North”
         start_date: "2015-04-23T10:00:00Z"
         end_date: "2016-05-26T23:00:00Z”
         instrument:
            model: “BBOBS1_1”
            serial_number: ”04"
            pressure_serial_number: "IP004"
            pressure_manufacturer: "IPGP"
            seismometer_serial_number: "138"
            seismometer_calibration_code: "1-399"
         locations:
            "00":
               <<: &LOC_DEFAULTS
               <<: &LOC_ACOUSTIC
               latitude: 37.31960
               longitude: -32.27909
               elevation: -1798
         non-standard:
            clock_correction_linear:
               <<: *LINEAR_CLOCK_DEFAULTS
               start_sync_reference: "2015-04-22T09:21:00Z”
               end_sync_reference: "2016-05-28T22:59:00.1843Z"
               end_sync_instrument: "2016-05-28T22:59:02Z"
      ”LSVNC”:
         ...
```

Limiting factor: lack of standards

- How to handle clock drift in data:
  - Three basic ideas, surprising level of discord
    - Correct the clock for each miniSEED record and indicate the correction used in the same header
    - Indicate the correction to use but do NOT apply it
    - Resample the data to the originally desired sampling rate
  - We can't make a standard software if no-one agrees on what it should create!
  - If you have an opinion (or alternative), please provide it at:
    - https://goo.gl/forms/zcDEAPj4n7kljjAt2
- How to specify OBS-specific metadata

Limiting factor: lack of standards

- How to put OBS-specific info into Station XML
  - How the clock correction was determined and implemented
  - How the instrument position was determined
    - Add a « Measurement method » attribute to the «uncertaintyDouble » type
  - How (if) the horizontal component orientation was determined
  - Options:
    - Use <Comment> tags,
      - Unstructured, text will often be very long
    - Make new <CommentList> tag:
      - Subject:
      - Comments (list)
    - Make new specific tags:
      - <Timebase>, <Horizontal_Orientation>, <Localization>, ...
  - Also may add a double-precision « sampling rate » blockette to miniSEED which will allow the true sampling rate to be specified
    - Standard OBS clocks have ~1e-8 drift (~1 second / year)
    - Chip-scale atomic clocks have ~1e-9.5 drift (~0.02 seconds/year)
    - Current single precision sampling rate blockette (B100) has 1e-7 precision (22-bit mantissa)

OBS-specific data tools (outside EPOS)

- Orientation
  - Several software tools have been developed, based on earthquake arrivals, whale songs and ship tracks
- Noise removal
  - Software exists to remove noise from vertical channel at low (<0.1 Hz frequencies), can reduce noise levels to broadband land station levels.
- Clock correction verification
  - Can be done using noise correlation between stations, changes in earthquake location time residuals over the course of the experiment

- These tools are generally written/used by individual researchers. By making standard versions that work on data center data publically available, these three significant problems can be greatly reduced.
  - OBSIP already makes one orientation code available online:
  - http://www.obsip.org/data/obs-horizontal-orientation/

Validation services (with CNRS engineer)

- Before send to data center
- To instrument park(s)
  - Station Instrument responses
  - (sub)Network Data availability
- To principal scientist
  - Network data availability
  - Event Waveforms
  - Station PPSDs
  - Network map

Other concerned groups

- FDSN
  - WG5 (mobile stations) working to develop standards
- OBSIP (US)
  - Have made most available OBS data
  - Their decisions have become *de facto* standards
  - Still room for improvement
- ENVRIplus
- EMSO
- Non-Europe and non-US countries
  - Taiwan has expressed interest
  - Japan has large fleet of OBSs and much data
  - China is developing an OBS fleet

Thank you!