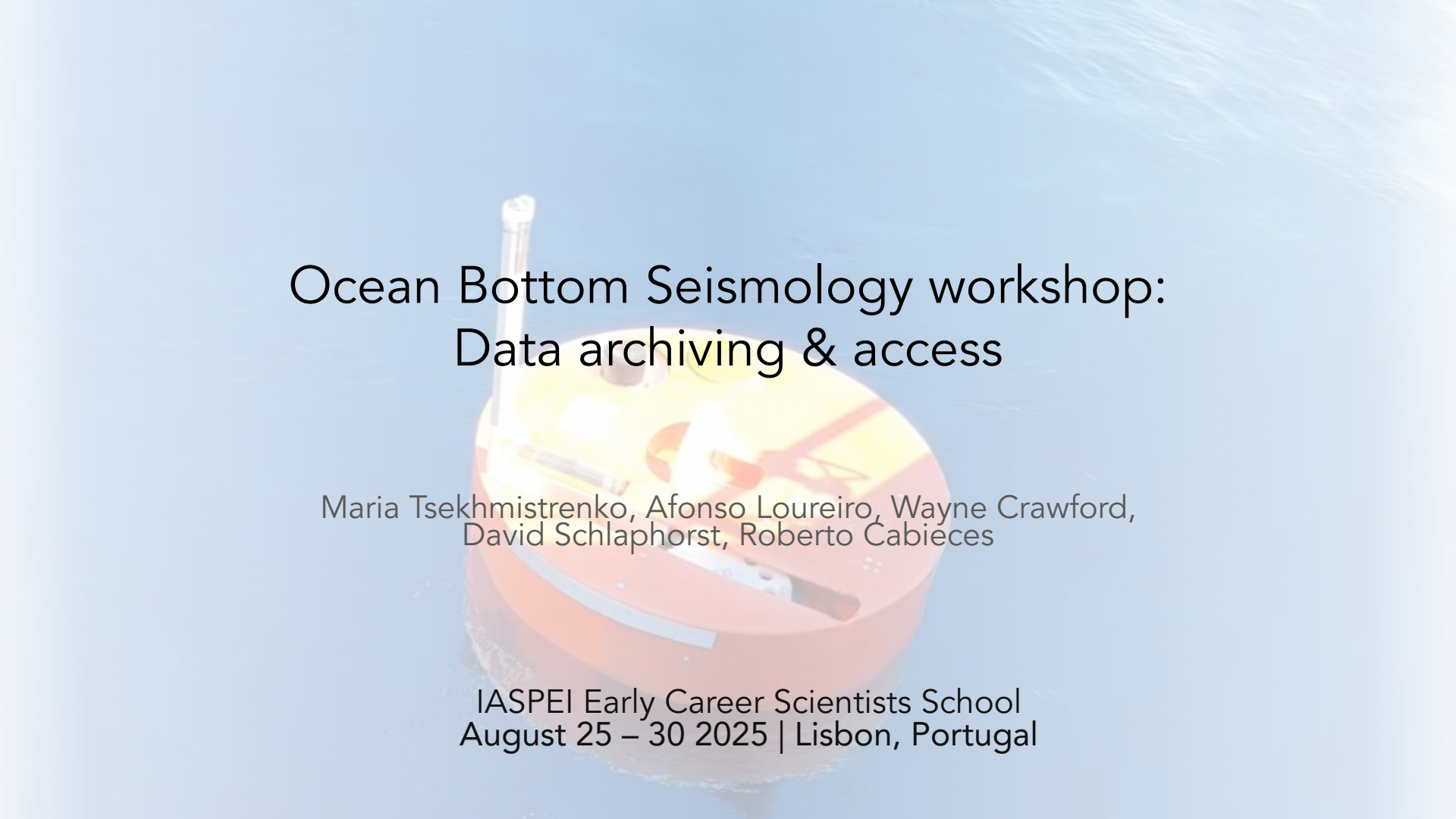


Ocean Bottom Seismology workshop: Data archiving & access

Maria Tsekhmistrenko, Afonso Loureiro, Wayne Crawford, David Schlaphorst, Roberto Cabieces

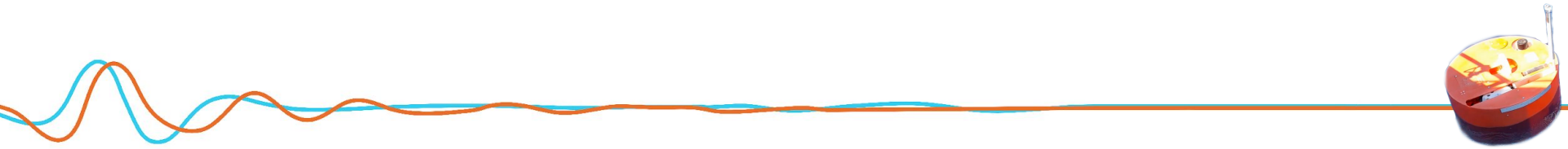
IASPEI Early Career Scientists School
August 25 – 30 2025 | Lisbon, Portugal

Data archiving

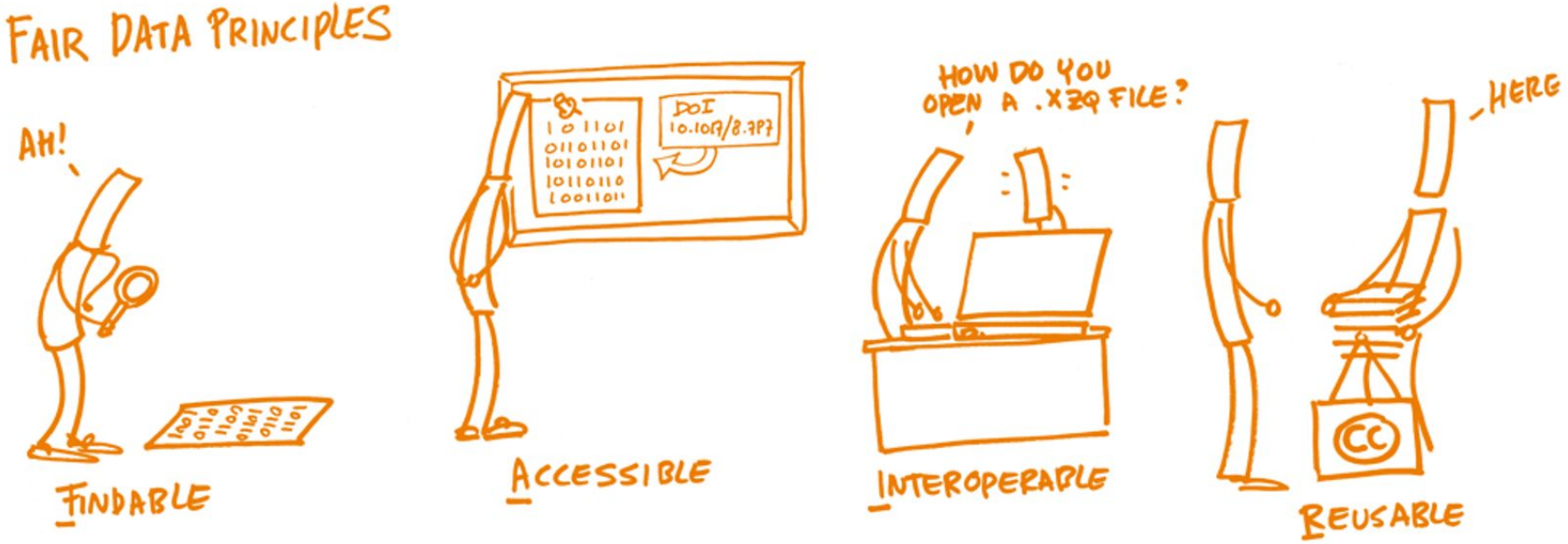
Slides provided by Laura Hillmann
Geofon_dc@gfz-Potsdam.de

Why upload the data at a data center?

- Easy distribution of data within the project and team members for the given embargo period
- Archiving & Back up is secured
- Citation via DOI (assigned by data center)
- Fairness Factor: “Give-and-take”



Why upload the data at a data center?



DOI for seismic networks

- Seismic Networks have a standard (FDSN) procedure to mint DOIs
  - Metadata includes
  - Author
  - Year
  - Title
  - Description
  - …

MATHEMATICAL GEOPHYSICS | Opinion

Why Seismic Networks Need Digital Object Identifiers

In a move to give credit where it's due, the International Federation of Digital Seismograph Networks will link digital object identifiers to data from seismic networks and project deployments.

A new system of digital object identifiers will bring long-overdue recognition to scientists providing data through seismic networks. Credit: Peter Evans

By Peter L. Evans, Angelo Strollo, Adam Clark, Tim Ahern, Robert Newman, John F. Clinton, Helle Pedersen, and Catherine Pequegnat · 8 October 2015

https://eos.org/opinions/why-seismic-networks-need-digital-object-identifiers

DOI for seismic networks

- Citing is based on network code
  - Journals have ‘Data Section’
  - If the network has a DOI: Refer to the data as you would to a normal scientific manuscript, and include the reference in the normal list of references.
  - Network, citation and DOI information can be found at http://www.fdsn.org/networks/
- Example: RHUM-RUM und UPFLOW have a DOI for their data:
  - https://www.fdsn.org/networks/detail/YV_2011/
  - https://www.fdsn.org/networks/detail/8J_2021/

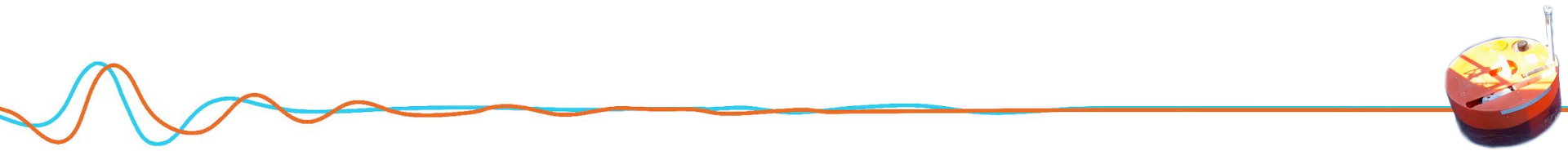
How to archive?

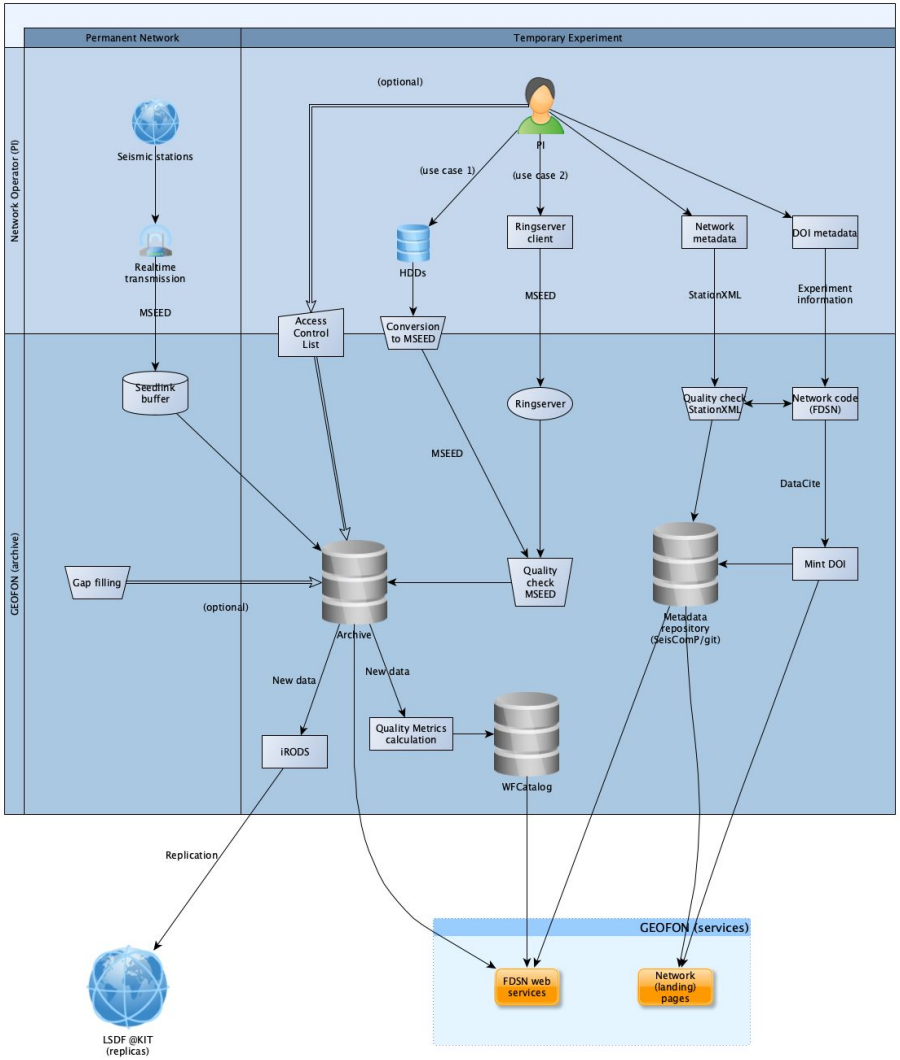
Preparation of:

- Data and
- Metadata

How and what to prepare, may depend on the data center.

➢ Thus ask them.





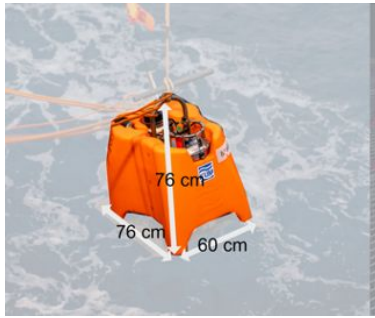
How to archive?

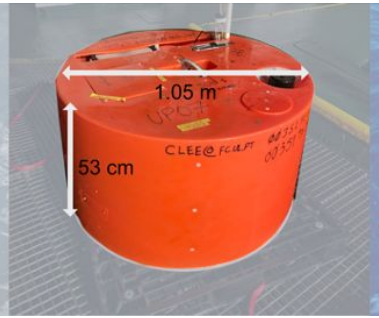
https://geofon.gfz.de/redmine/projects/redmine/wiki/Pub

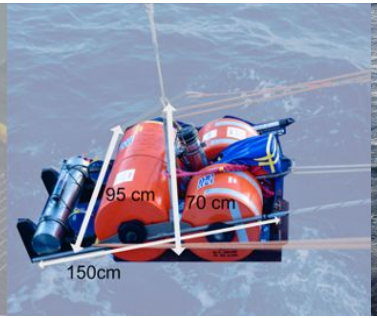
https://geofon.gfz.de/redmine/projects/redmine/wiki/GIPP

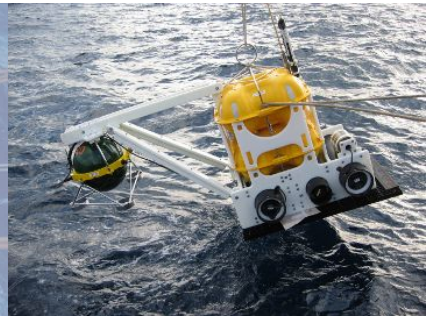
Ocean Bottom Sensors peculiarities

- Position uncertainties
- Unknown orientation
- No GPS on the ground
- time correction needed
- Less common instruments (self development)





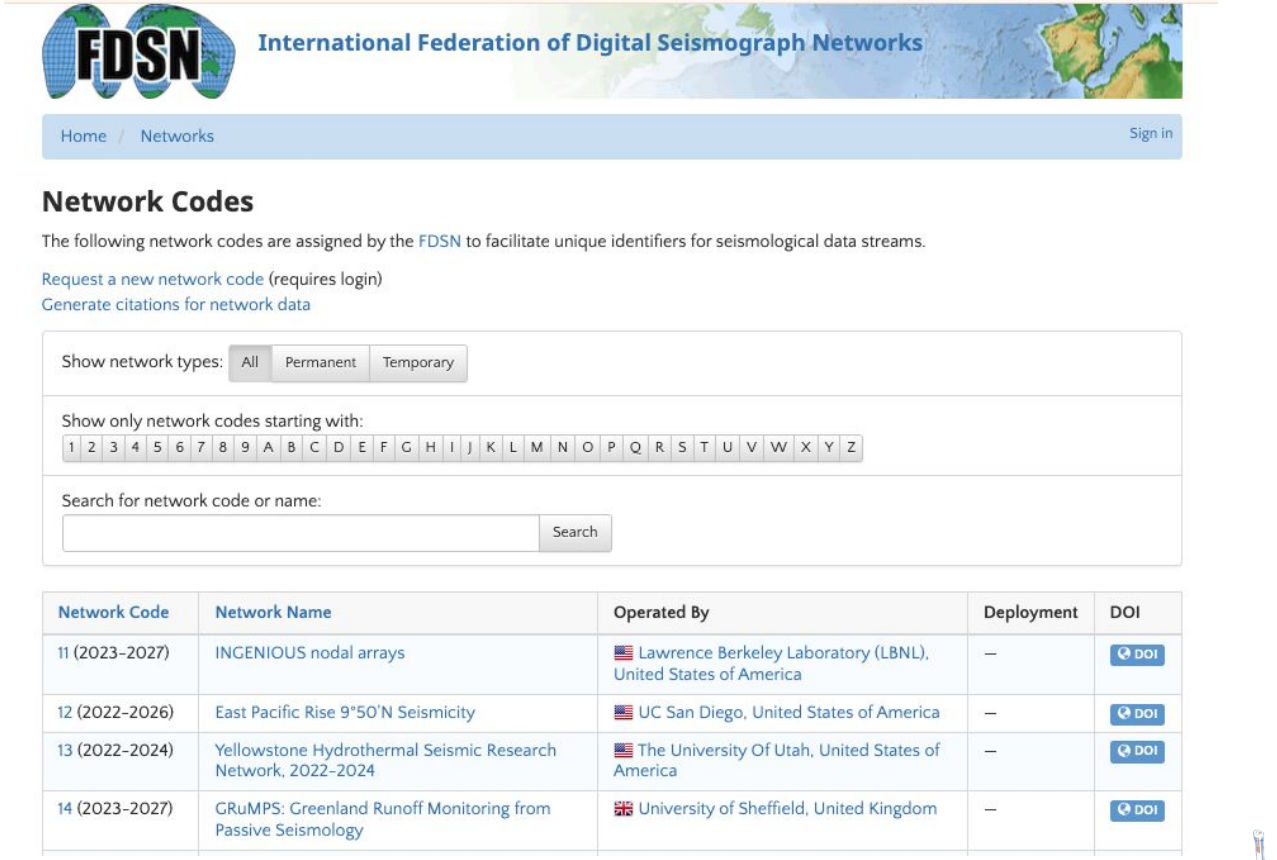




How to archive?

Step 1: Network Code

- Request a network code at FDSN



How to archive?

Step 2: Data Preparation

- header containing the final network code (assigned by FDSN)
- header containing appropriate station code, channel naming and location code
- blocksize (512 or 4096)
- ByteOrder (e.g. BigEndian)
- Compression: Steim1 or Steim2

What is…blocksize?

Think of a seismic data file as being split into blocks.

Common sizes: 512 bytes or 4096 bytes.

- Larger blocks (4096) → more efficient for modern systems and reduces overhead.
- Smaller blocks (512) → historically used, still common for compatibility.

Analogy: Imagine writing a book. You can decide whether each page holds 512 characters or 4096 characters. Bigger pages mean fewer page breaks, but you need to read/write more at once.

What is…ByteOrder?

Computers store numbers as a sequence of bytes. The order of these bytes can differ:

- Big Endian → the most significant byte comes first.
- Little Endian → the least significant byte comes first.

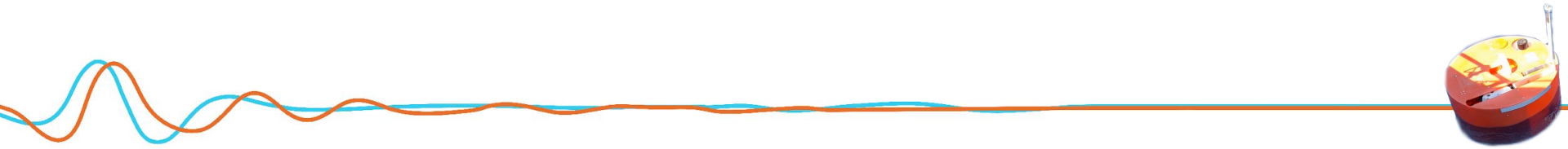
Which one is used depends on the computer system or file format convention.

Analogy: Writing the number 1234.

Big Endian: write it as 1-2-3-4.

Little Endian: write it as 4-3-2-1.

Both represent the same number, but you need to know the order to interpret it correctly.



What is…Steim1 and Steim2?

It comes from Dr. Steim, who developed the Steim 1 and Steim 2 compression algorithms in the late 1980s.

Seismic data takes up a lot of space because it's recorded at high rates over long times.

To save space, MiniSEED often uses Steim compression:

- Steim1 → simpler, slightly less efficient, very compatible.
- Steim2 → more efficient (better compression), but slightly more complex.

Both methods work by storing the differences between consecutive samples instead of the raw values.

Analogy: Instead of writing down every GPS position on a walk, you just note the steps you take from the previous point. This makes the data smaller but still allows full reconstruction.

Metadata

- Network code: 2 characters (assigned by FDSN)
- Station code: up to 5 characters
- Location code : 0 or 2 characters
- Channel codes: 3 characters (depending on sampling rate and seed conventions)

Instruments (Digitizer, Sensor, …) incl. serial numbers

Sampling rates

Orientation (XYZH, 12ZH, 1234, etc.)

Response Files

Example:

8J.UP05.00.HHZ

8J.UP12..CHZ

Information from the deployment sheet!

How to archive?

Step 3: Send Metadata

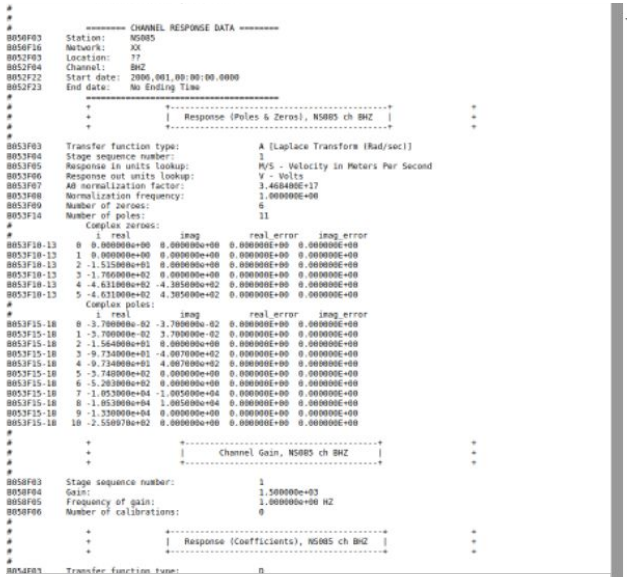
- Network code is assigned by FDSN
- Title of the experiment
- Start year
- End year
- Short abstract (max 200 chars)
- Estimated size (in Megabytes)
- Operating institution(s)
- PI(s) (ideally with ORCID iD)
- Funder(s)
- Deployment area
- Will the data be embargoed?
- Expected release date
- License (now or at end of the embargo

Most of the information will be reused for mining the DOI.
Some data centers offer to request the network code for you.

How to archive?

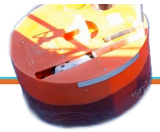
Send as:

- StationXML
- Resp(onse) files
- Dataless Seed



```
#
#
#          ======== CHANNEL RESPONSE DATA ========
B050F03     Station:     NS085
B050F16     Network:     XX
B052F03     Location:    ??
B052F04     Channel:     BHZ
B052F22     Start date:  2000,001,00:00:00.0000
B052F23     End date:    No Ending Time
#           ========================================
#            +                 +-------------------------------------+             +
#            +                 |  Response (Poles & Zeros), NS085 ch BHZ  |             +
#            +                 +-------------------------------------+             +
#
B053F03     Transfer function type:               A [Laplace Transform (Rad/sec)]
B053F04     Stage sequence number:                1
B053F05     Response in units lookup:             M/S - Velocity in Meters Per Second
B053F06     Response out units lookup:            V - Volts
B053F07     A0 normalization factor:              3.468400E+17
B053F08     Normalization frequency:              1.000000E+00
B053F09     Number of zeroes:                     6
B053F14     Number of poles:                      11
#           Complex zeroes:
#             i  real          imag          real_error    imag_error
B053F10-13    0  0.000000e+00  0.000000e+00  0.000000E+00  0.000000E+00
B053F10-13    1  0.000000e+00  0.000000e+00  0.000000E+00  0.000000E+00
B053F10-13    2 -1.515000e+01  0.000000e+00  0.000000E+00  0.000000E+00
B053F10-13    3 -1.766000e+02  0.000000e+00  0.000000E+00  0.000000E+00
B053F10-13    4 -4.631000e+02 -4.305000e+02  0.000000E+00  0.000000E+00
B053F10-13    5 -4.631000e+02  4.305000e+02  0.000000E+00  0.000000E+00
#           Complex poles:
#             i  real          imag          real_error    imag_error
B053F15-18    0 -3.700000e-02 -3.700000e-02  0.000000E+00  0.000000E+00
B053F15-18    1 -3.700000e-02  3.700000e-02  0.000000E+00  0.000000E+00
B053F15-18    2 -1.564000e+01  0.000000e+00  0.000000E+00  0.000000E+00
B053F15-18    3 -9.734000e+01 -4.007000e+02  0.000000E+00  0.000000E+00
B053F15-18    4 -9.734000e+01  4.007000e+02  0.000000E+00  0.000000E+00
B053F15-18    5 -5.748000e+02  0.000000e+00  0.000000E+00  0.000000E+00
B053F15-18    6 -5.203000e+02  0.000000e+00  0.000000E+00  0.000000E+00
B053F15-18    7 -1.053000e+04 -1.005000e+04  0.000000E+00  0.000000E+00
B053F15-18    8 -1.053000e+04  1.005000e+04  0.000000E+00  0.000000E+00
B053F15-18    9 -1.330000e+04  0.000000e+00  0.000000E+00  0.000000E+00
B053F15-18   10 -2.550970e+02  0.000000e+00  0.000000E+00  0.000000E+00
#
#            +                 +-------------------------------------+             +
#            +                 |      Channel Gain, NS085 ch BHZ     |             +
#            +                 +-------------------------------------+             +
#
B058F03     Stage sequence number:                1
B058F04     Gain:                                 1.500000e+03
B058F05     Frequency of gain:                    1.000000e+00 HZ
B058F06     Number of calibrations:               0
#
#            +                 +-------------------------------------+             +
#            +                 |  Response (Coefficients), NS085 ch BHZ  |             +
#            +                 +-------------------------------------+             +
#
B054F03     Transfer function type:               D
```


```
-<FDSNStationXML schemaVersion="1.1">
  <Source>SeisComP</Source>
  <Sender>GFZ</Sender>
  <Created>2022-05-20T11:58:00.427604</Created>
 -<Network code="GE" startDate="1993-01-01T00:00:00" restrictedStatus="open">
    <Description>GEOFON Program, GFZ Potsdam, Germany</Description>
    <Identifier type="DOI">10.14470/TR560404</Identifier>
   -<Station code="RGN" startDate="1995-12-01T00:00:00" restrictedStatus="open">
      <Latitude>54.5477</Latitude>
      <Longitude>13.3214</Longitude>
      <Elevation>15</Elevation>
     -<Site>
        <Name>GRSN/GEOFON Station Ruegen, Germany</Name>
        <Country>Germany</Country>
      </Site>
      <CreationDate>1995-12-01T00:00:00</CreationDate>
     -<Channel code="HHZ" startDate="2014-05-02T12:48:00" restrictedStatus="open" locationCode="">
        <Latitude>54.5477</Latitude>
        <Longitude>13.3214</Longitude>
        <Elevation>15</Elevation>
        <Depth>2</Depth>
        <Azimuth>0</Azimuth>
        <Dip>-90</Dip>
        <SampleRate>100</SampleRate>
       -<SampleRateRatio>
          <NumberSamples>100</NumberSamples>
          <NumberSeconds>1</NumberSeconds>
        </SampleRateRatio>
        <ClockDrift>0</ClockDrift>
       -<Sensor resourceId="Sensor/20190409123644.652274.6">
          <Type>VBB</Type>
          <Description>GFZ:GE1993:STS-2/N/g=1500</Description>
          <Manufacturer>Streckeisen</Manufacturer>
          <Model>STS-2/N</Model>
        </Sensor>
       -<DataLogger resourceId="Datalogger/20190409123644.659328.35">
          <Type>EarthData PS6-SC</Type>
          <Description>GFZ:GE1993:PS6-SC/g=392000</Description>
          <Manufacturer>EarthData</Manufacturer>
          <Model>PS6-SC</Model>
        </DataLogger>
       -<Response>
         -<InstrumentSensitivity>
            <Value>588000000</Value>
            <Frequency>0.02</Frequency>
           -<InputUnits>
              <Name>M/S</Name>
            </InputUnits>
           -<OutputUnits>
              <Name>COUNTS</Name>
            </OutputUnits>
          </InstrumentSensitivity>
         -<Stage number="1">
```

How to archive?

Table containing the following information:

| StationCode | Place/Country | DataLogger Device Gain | Sensor SerialNumber | Latitude | Longitude | Elevation | SamplingRate | LocalDepth | AziChan1 | AziChan2 | StartDate | EndDate |

- Place / Country means nearest village or spot that helps to distinguish stations or otherwise be left empty
- AziCHan1/AziChan2 only necessary if instruments are not oriented, otherwise it's Z,N,E
- Start / End means start and end of the deployment of the station
- Device means the serial number of the instrument which gets more and more important because of the increasing use of persistent identifiers (PIDs) for instruments.

Data Transfer via Ringserver

Now you have to download and install the Ringserver Client software miniseed2dmc

- do Dry Runs (without server connection) for testing purposes, in order to find out if the data is recognized as miniSEED
- transfer of a small amount of data to see if all criteria have been met
- Call miniseed2dmc -v 139.17.3.77:port
  - port number being assigned by our data centre.
- Your data and metadata will be checked several times:
  - while preparing the metadata and data, parsers are checking the consistency of the generated StationXML
  - after archiving PSD files are computed
  - measures will be taken to make sure that data will stay untouched after final archival
- Errors and mistakes will always be corrected also much later, whenever they may be detected.

Access

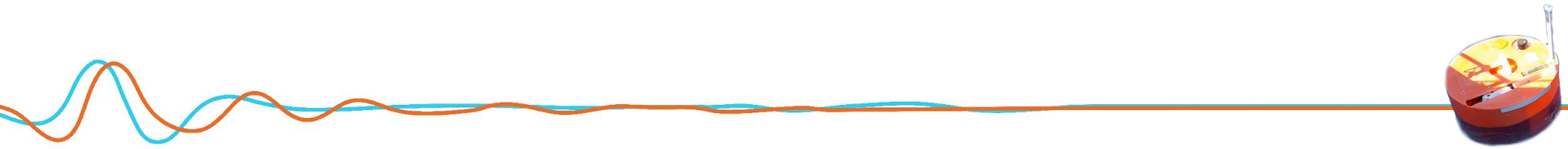
Slides provided by Laura Hillmann
Geofon_dc@gfz-Potsdam.de

How to download/access data

- URL
- Obspy clients for web services or obspyDMT
- WebDC3 (interactive portal)
- …

Difference between open and restricted data

- For restricted data access you need a token
- PI can give you access



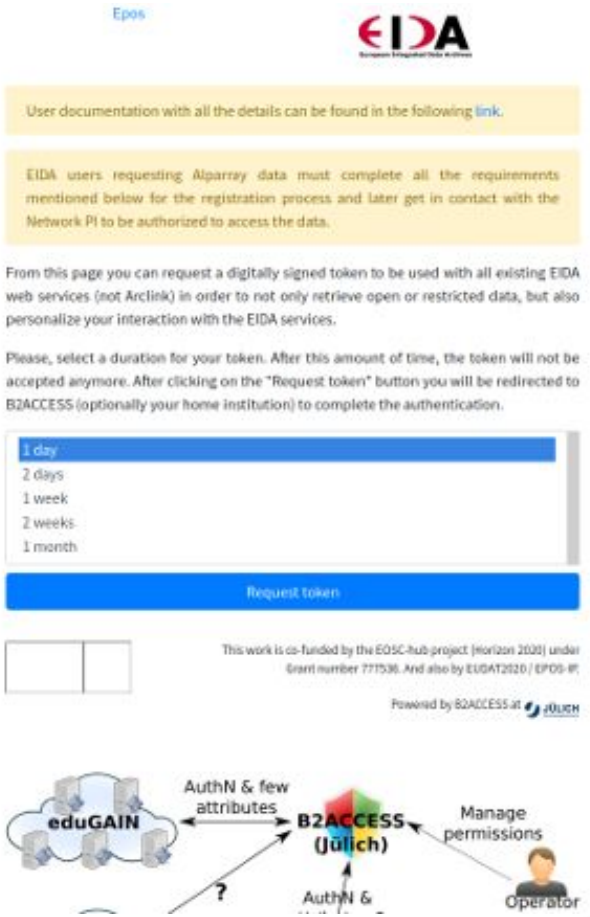
Token

For restricted data stored at an EIDA data center:

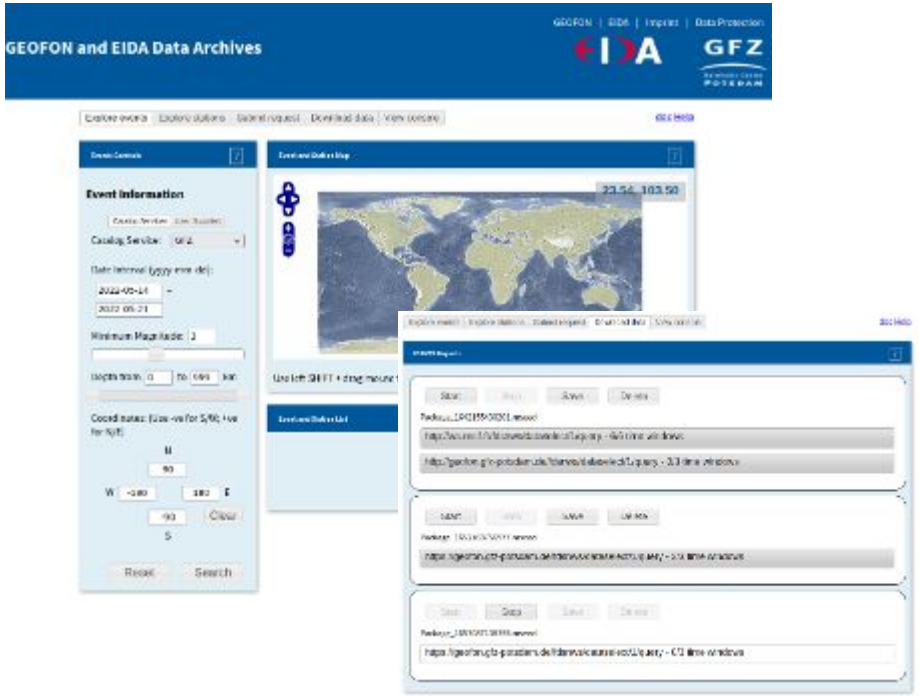
https://geofon.gfz-potsdam.de/eas/

The token is kept secret on the user's side.

- Example for FDSNWS with password: https://geofon.gfz-potsdam.de/waveform/archive/auth/auth-example.php



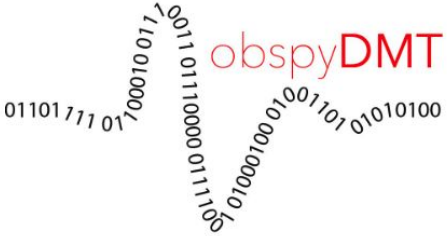
WebDC3



http://eida.gfz-potsdam.de/webdc3/

- WebDC3 is an interactive portal to browse and request tailor-made datasets, explore events from different catalogs using standard seismological services.
- Adopted by the European Integrated data Archive (EIDA). Currently run at seismological data centers in Germany, the Netherlands, Switzerland, Romania, Brazil, Indonesia and more.

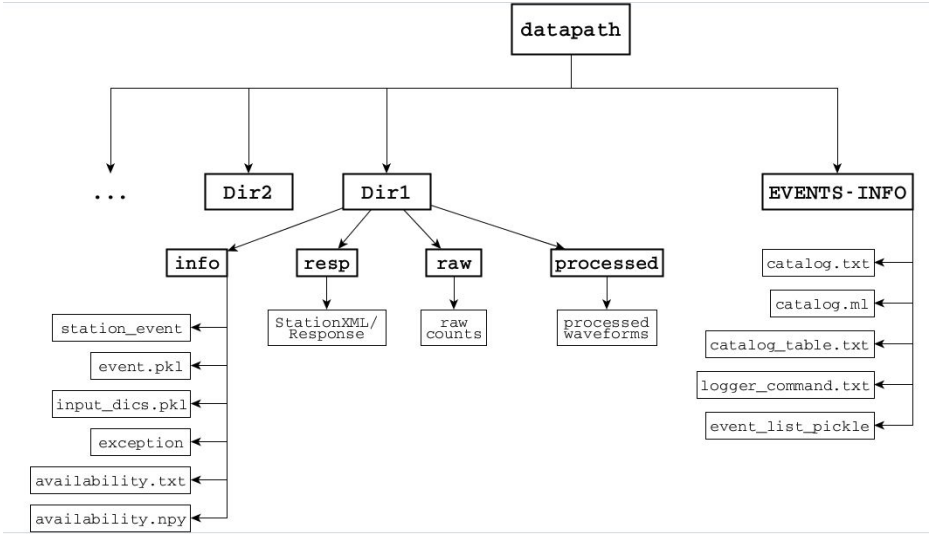
obspyDMT



A Python Toolbox for Retrieving, Processing and Management of Seismological Datasets

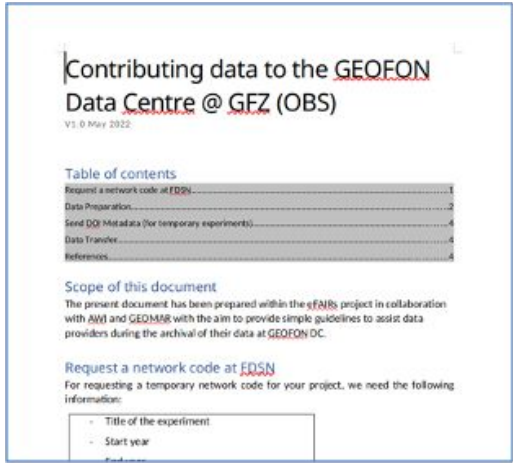
pypi v2.2.11 | License LGPL v3 | Integration Tests no status

obspyDMT (obspy Data Management Tool) is a tool for retrieving, processing and management of seismological datasets in a fully automatic way. Written in the Python programming language (Python 2 and 3 compatible), it can be used as a stand-alone command-line tool (requiring no knowledge of Python) or can be integrated as a module with other Python codes.



Work done by Laura Hillmann et al.

- Worked on joint project with AWI, GEOMAR and Wayne Crawford
- involvement in metadata standards for OBS data
- eFAIRs - enhancing FAIRness in seismological data management



Contributing data to the GEOFON Data Centre @ GFZ (OBS)

V1.0 May 2022

Table of contents

Scope of this document

The present document has been prepared within the eFAIRs project in collaboration with AWI and GEOMAR with the aim to provide simple guidelines to assist data providers during the archival of their data at GEOFON DC.

Request a network code at FDSN

For requesting a temporary network code for your project, we need the following information:

- Title of the experiment
- Start year

eFAIRs - enhancing FAIRness in seismological data management

Objectives

- automate the integration of instrument PIDs (Persistent Identifiers) in the standard metadata generation workflow
- adoption of standard and interoperable vocabularies in the DOI and EPIC Identifiers metadata
- develop tailored metadata consistency/quality checks
- achieve full integration of OBS data sets in the routine seismological data management workflow for long term archival
- foster adoption and usage of these fully FAIR data management policies within the community