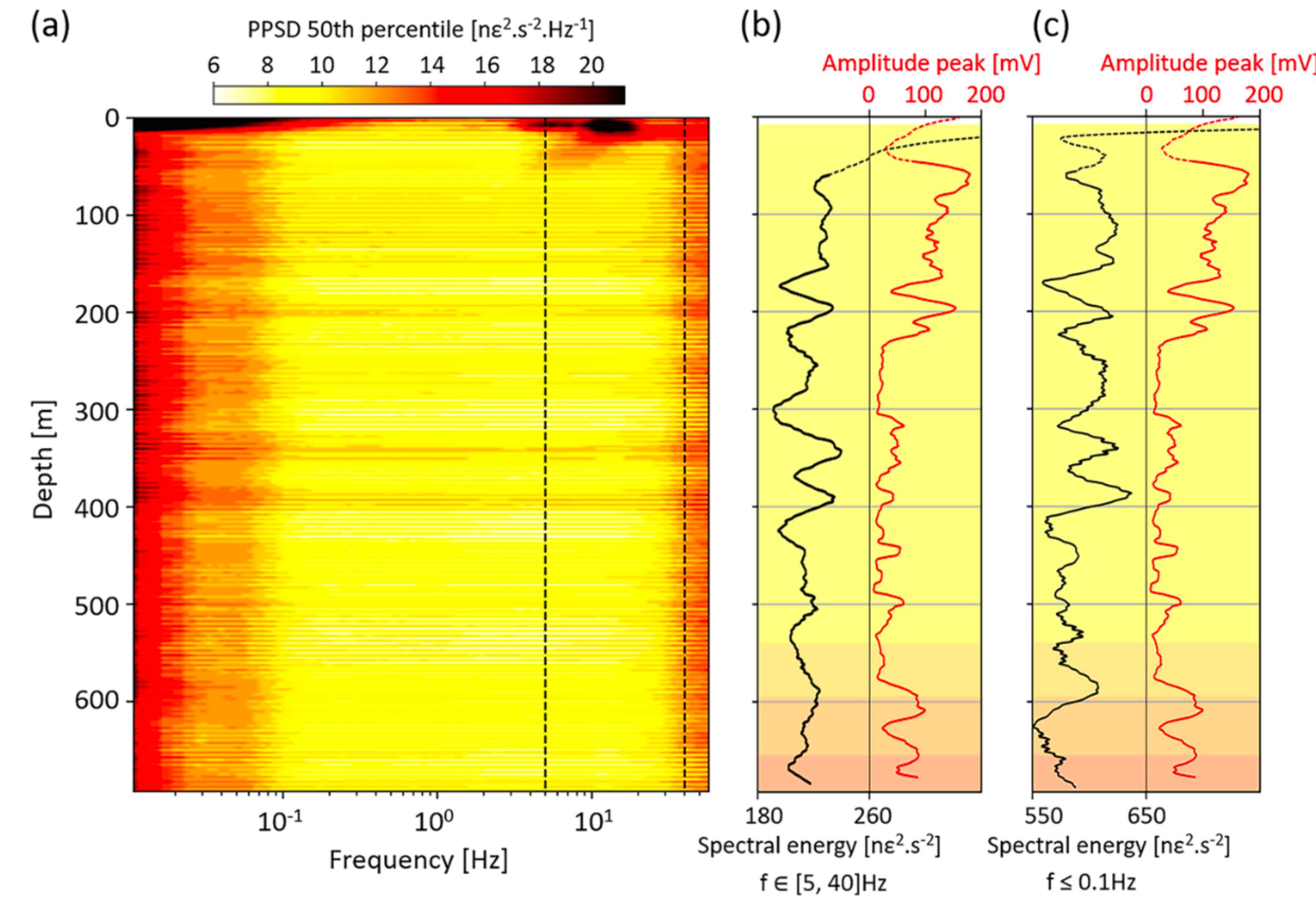
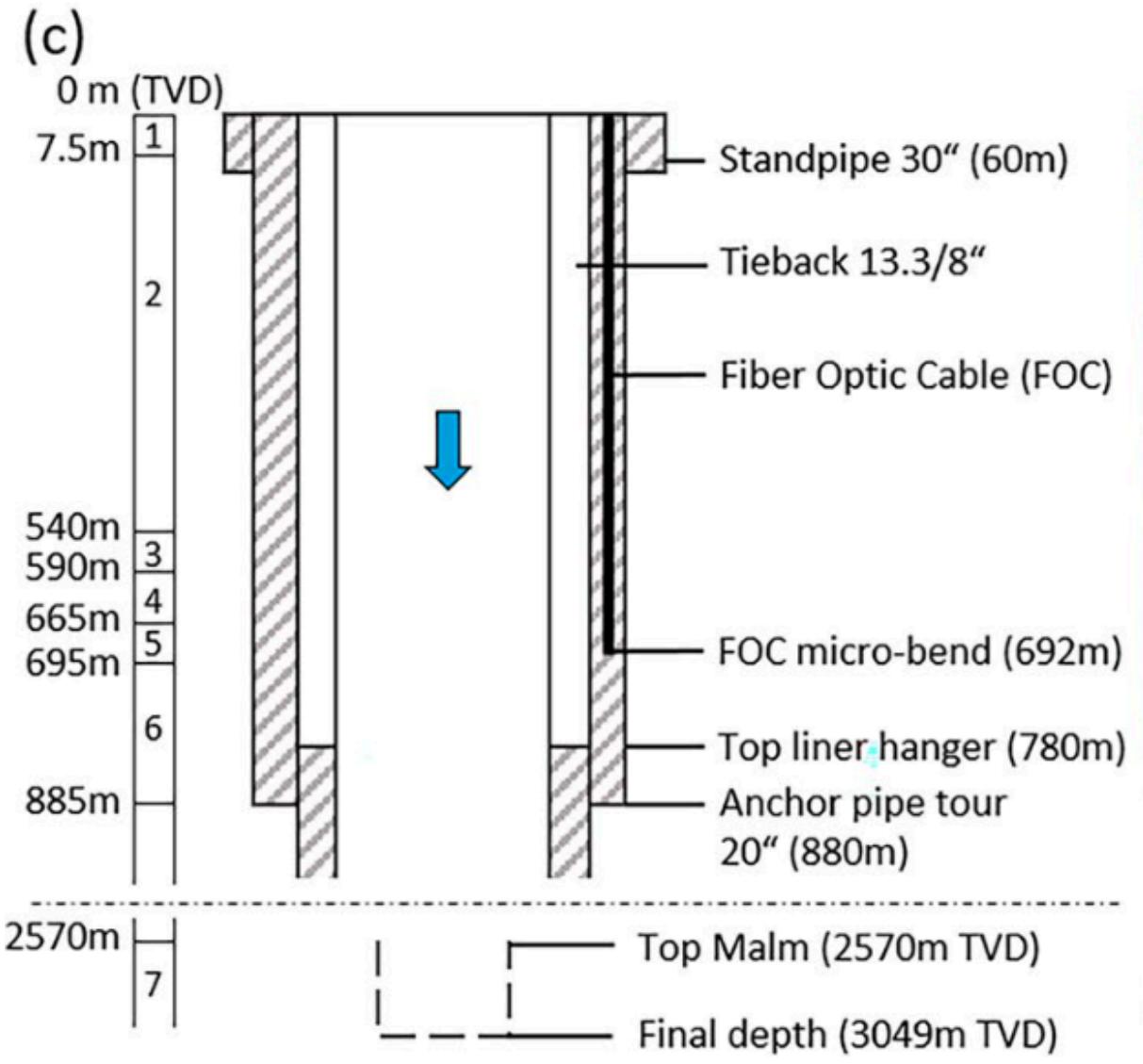
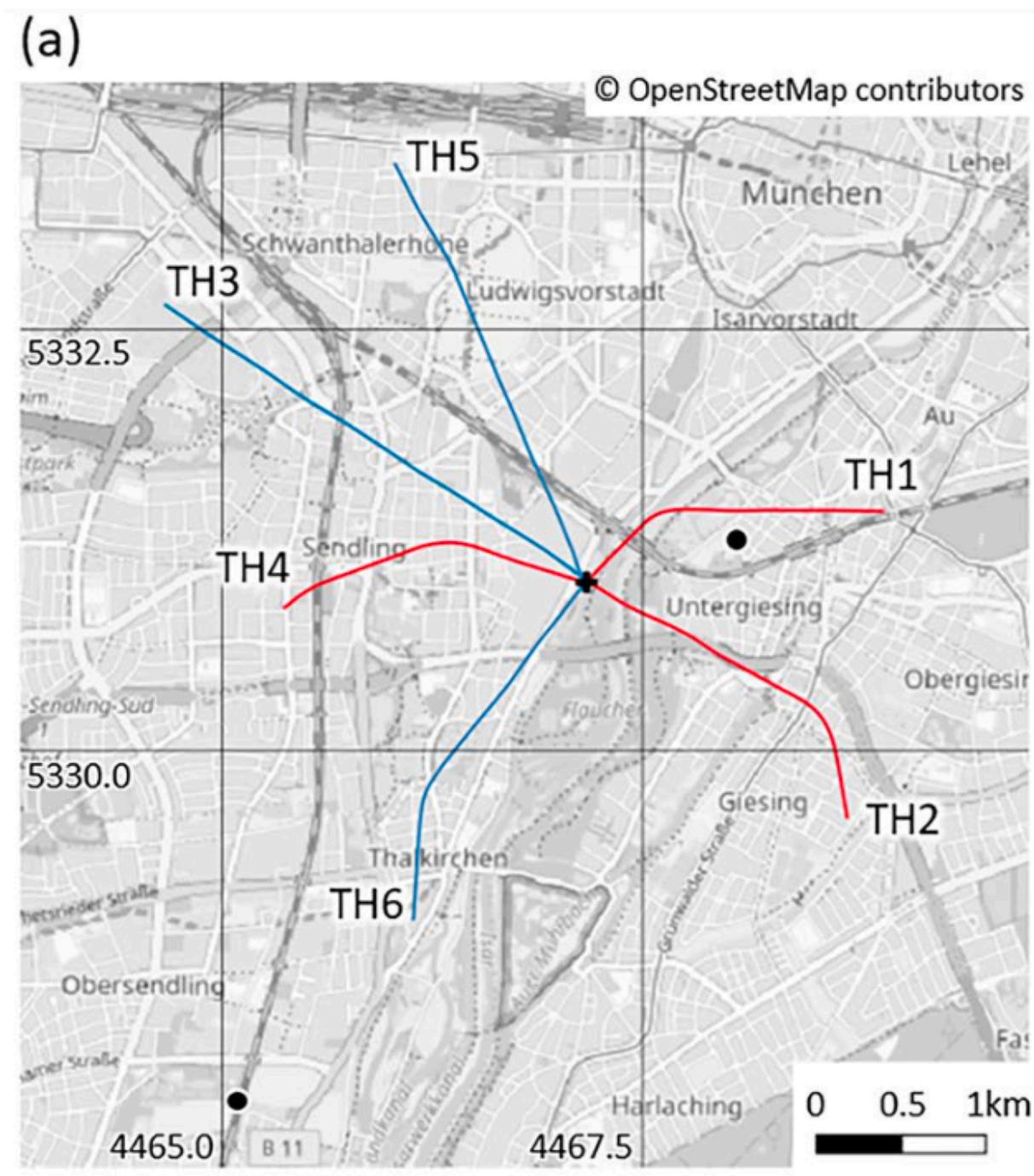


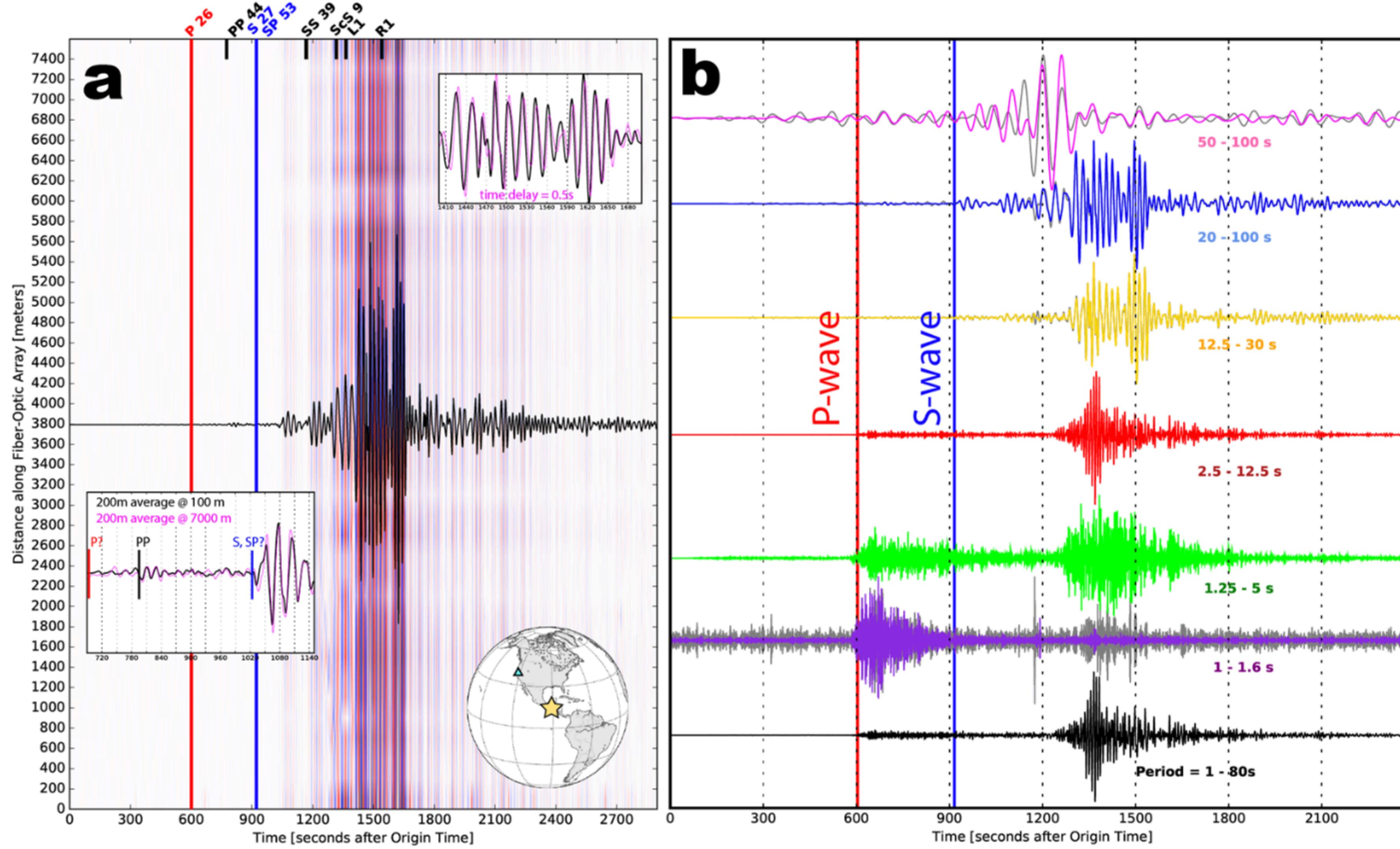
# The challenge and current status of distributed acoustic sensing (DAS) data management

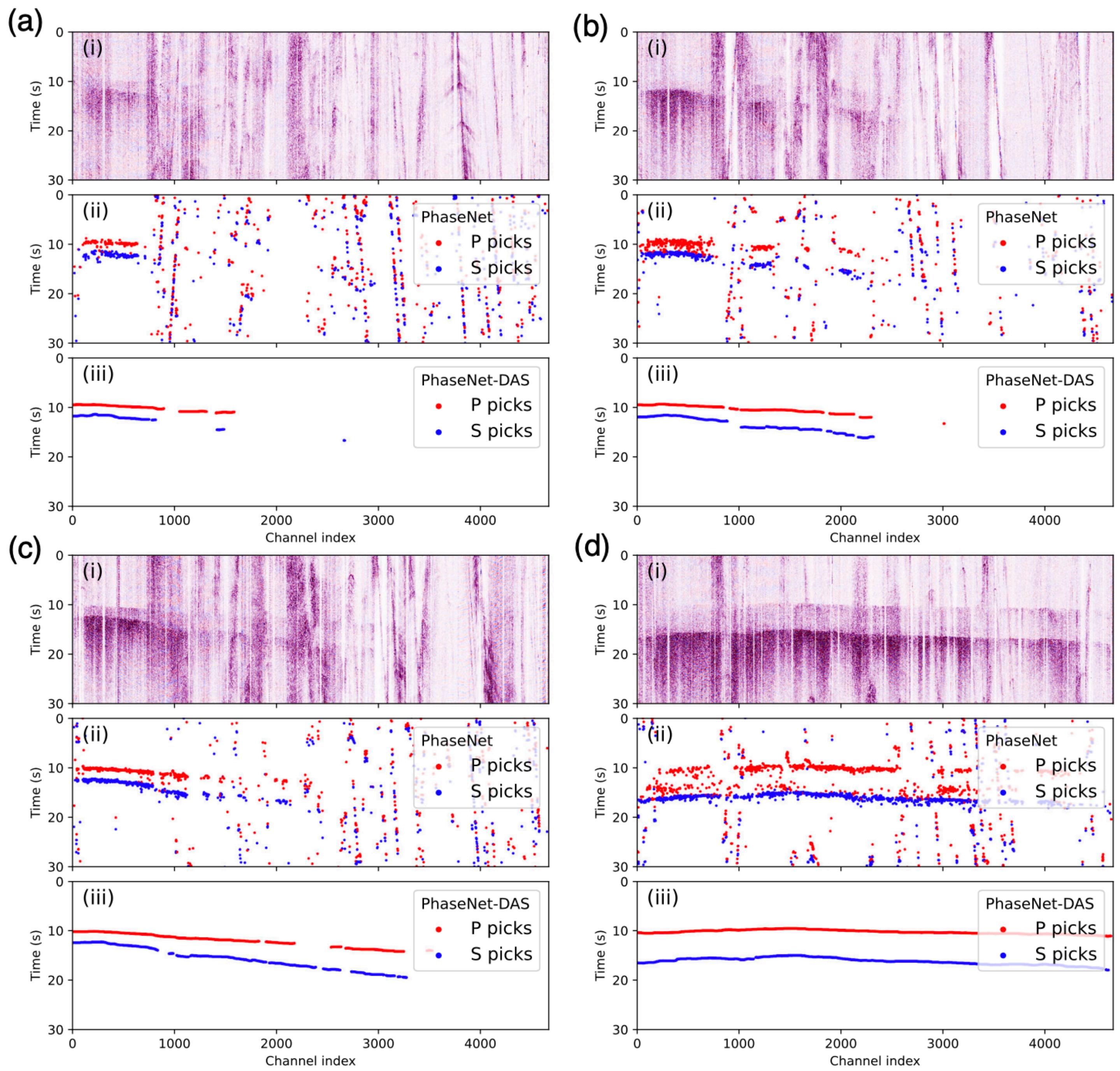
Chin-Shang Ku  
2024/01/12 @DAS workshop

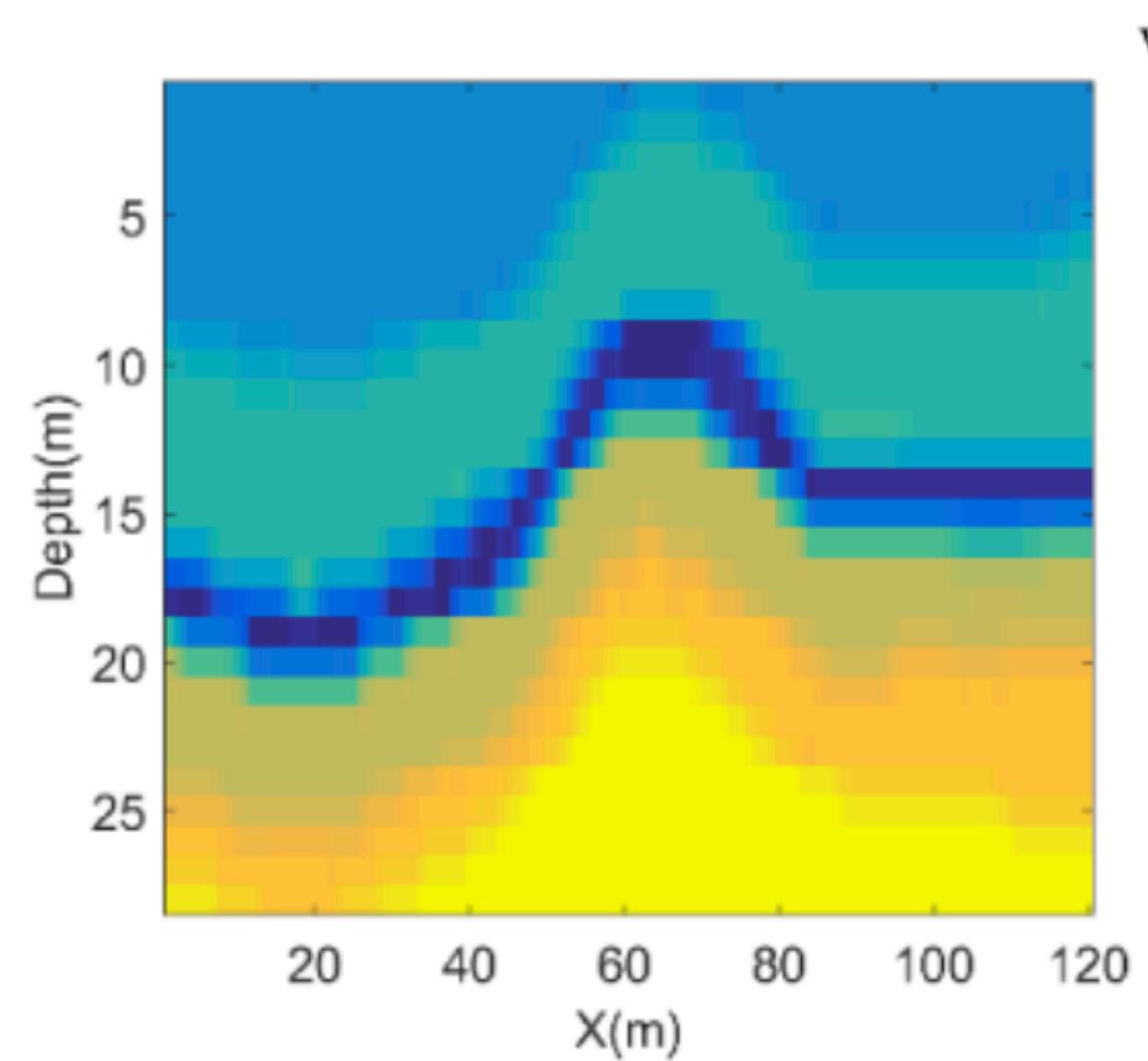
# The advantage of DAS and its applications

- The oil and gas industry
- Geothermal energy development
- Carbon capture and storage
- Near-surface geophysics
- Engineering infrastructure monitoring
- Environmental research
- Seismic monitoring
- ...

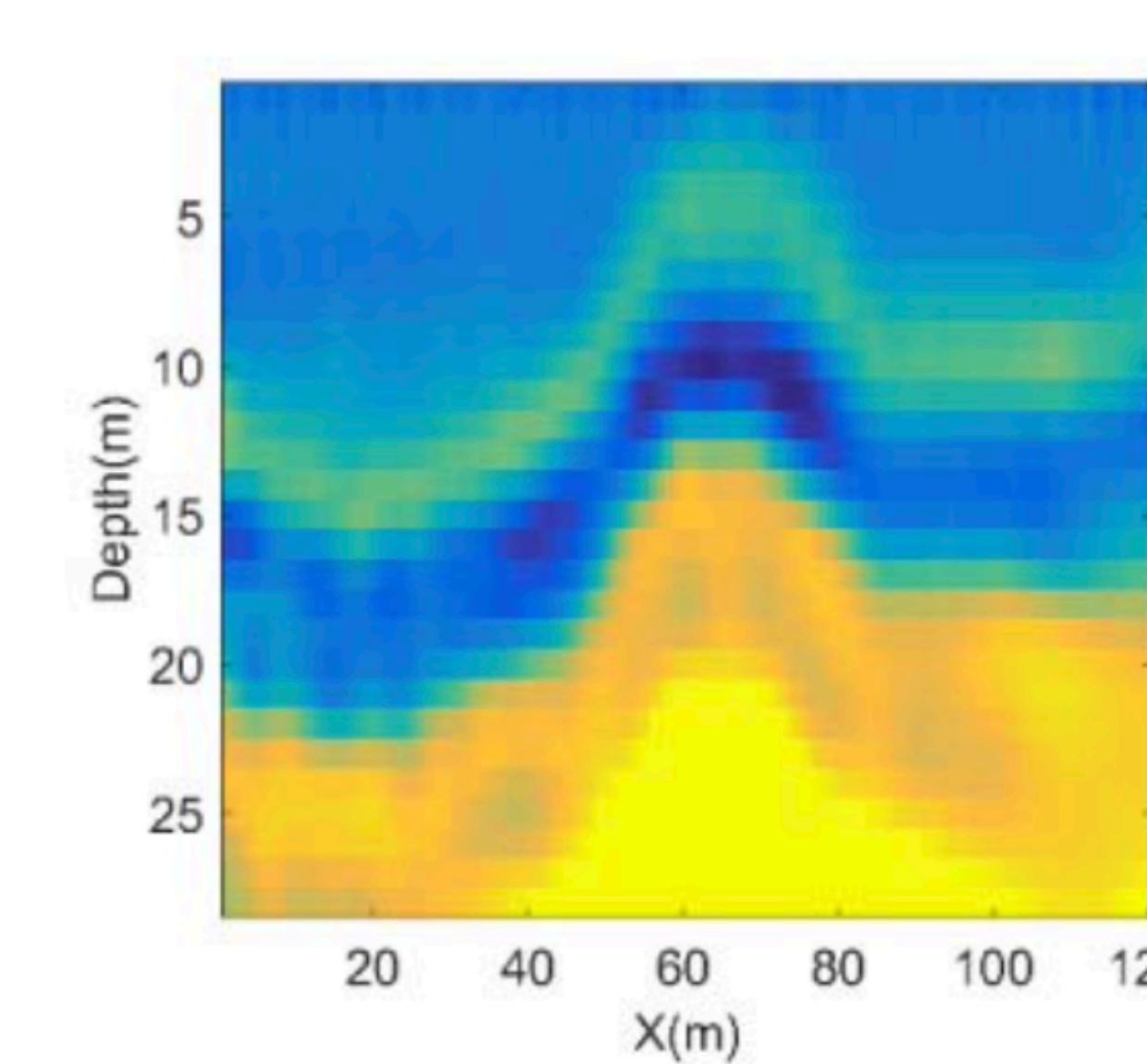




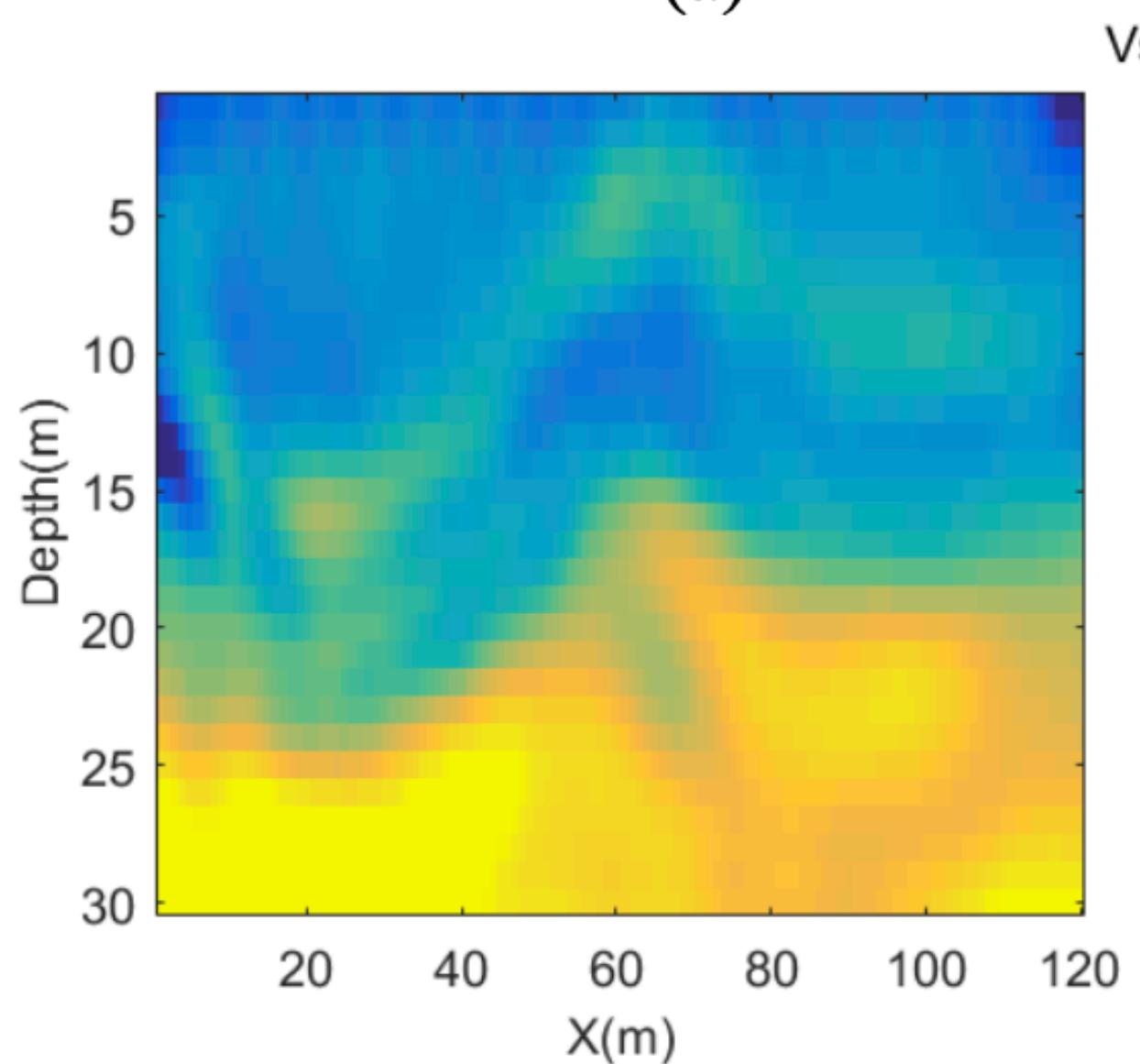




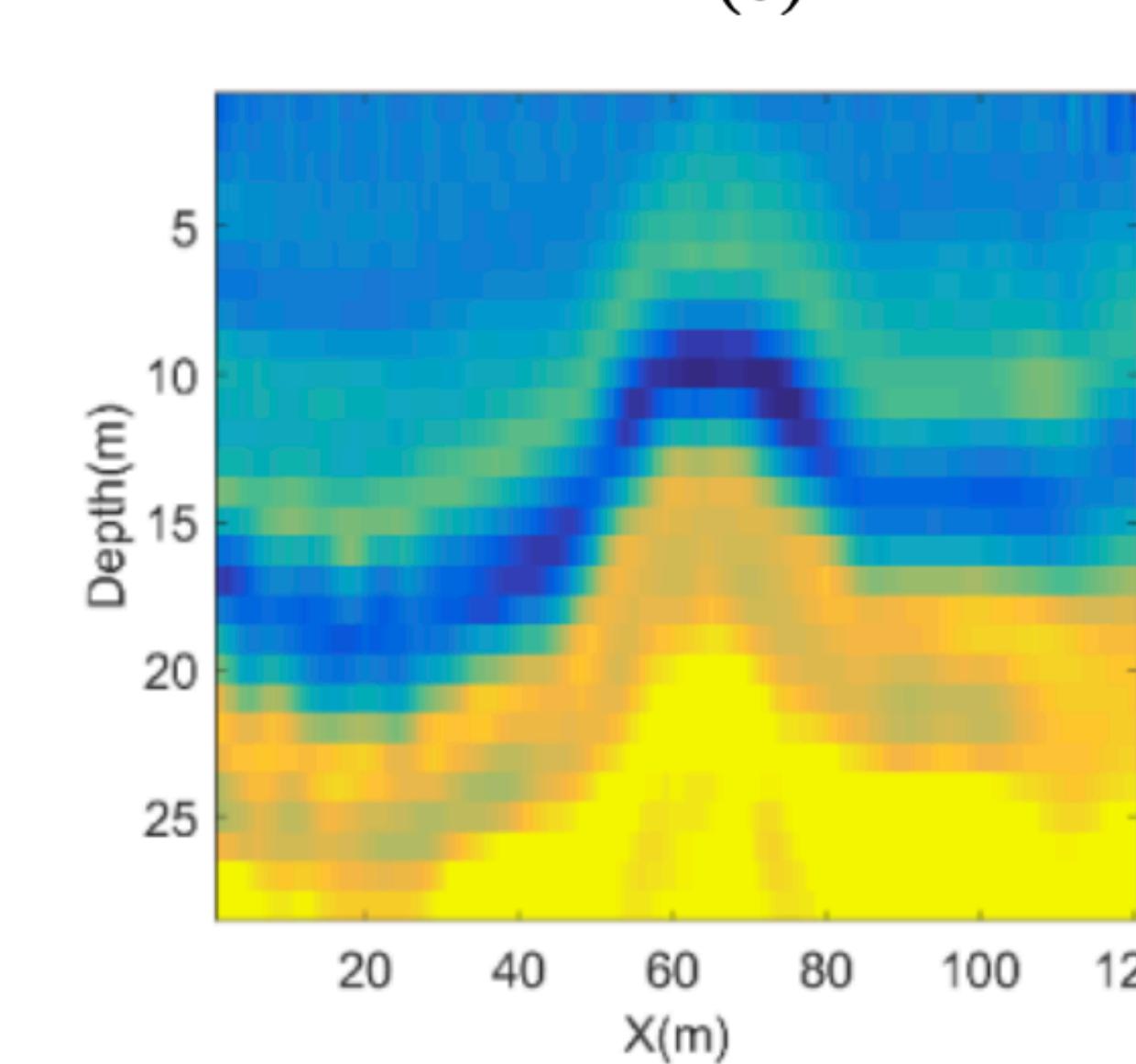
(a)



(b)

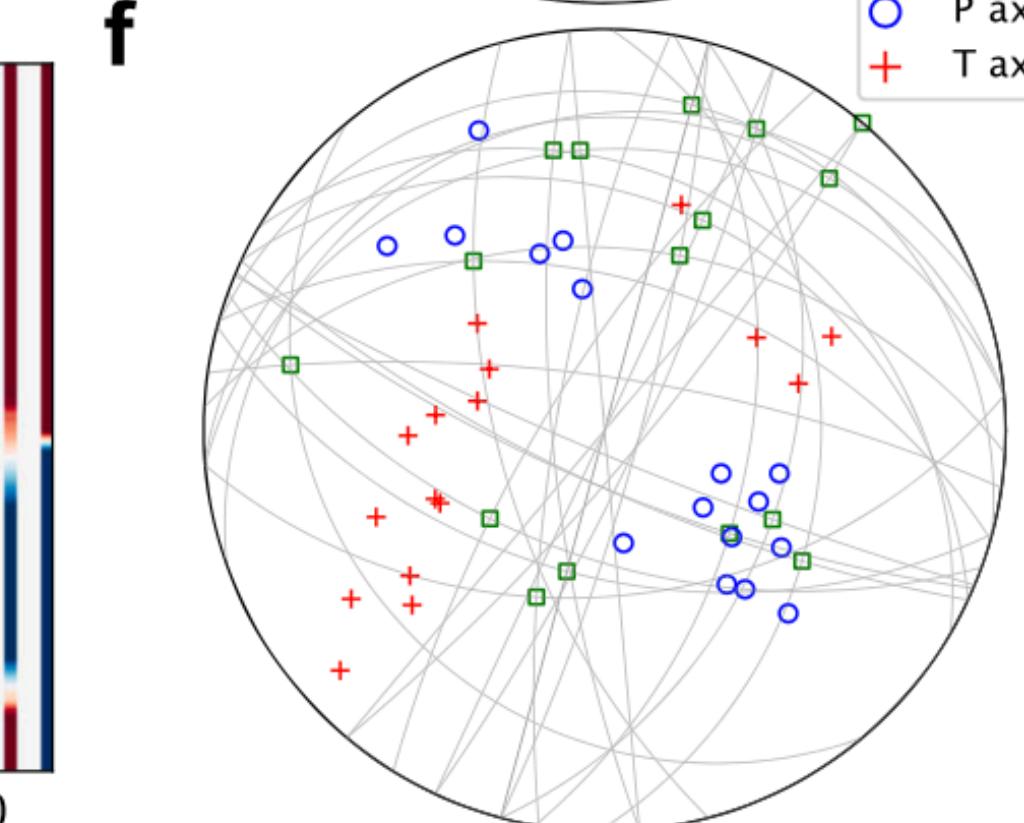
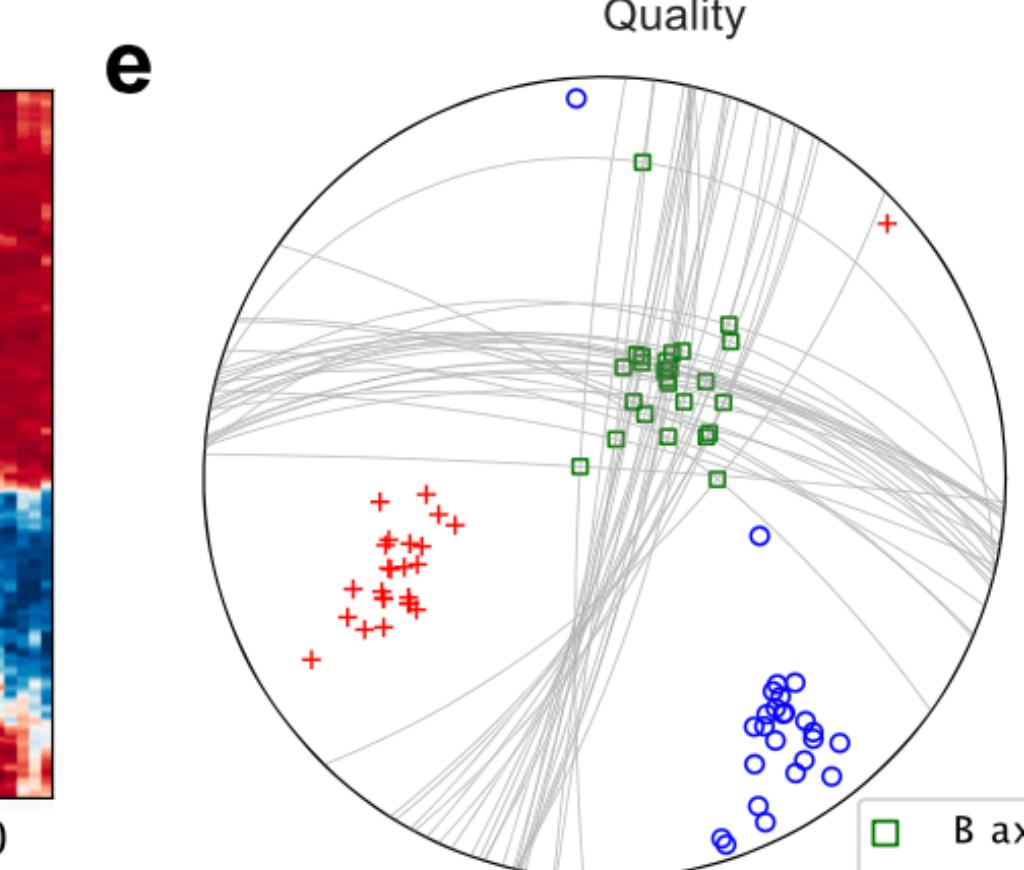
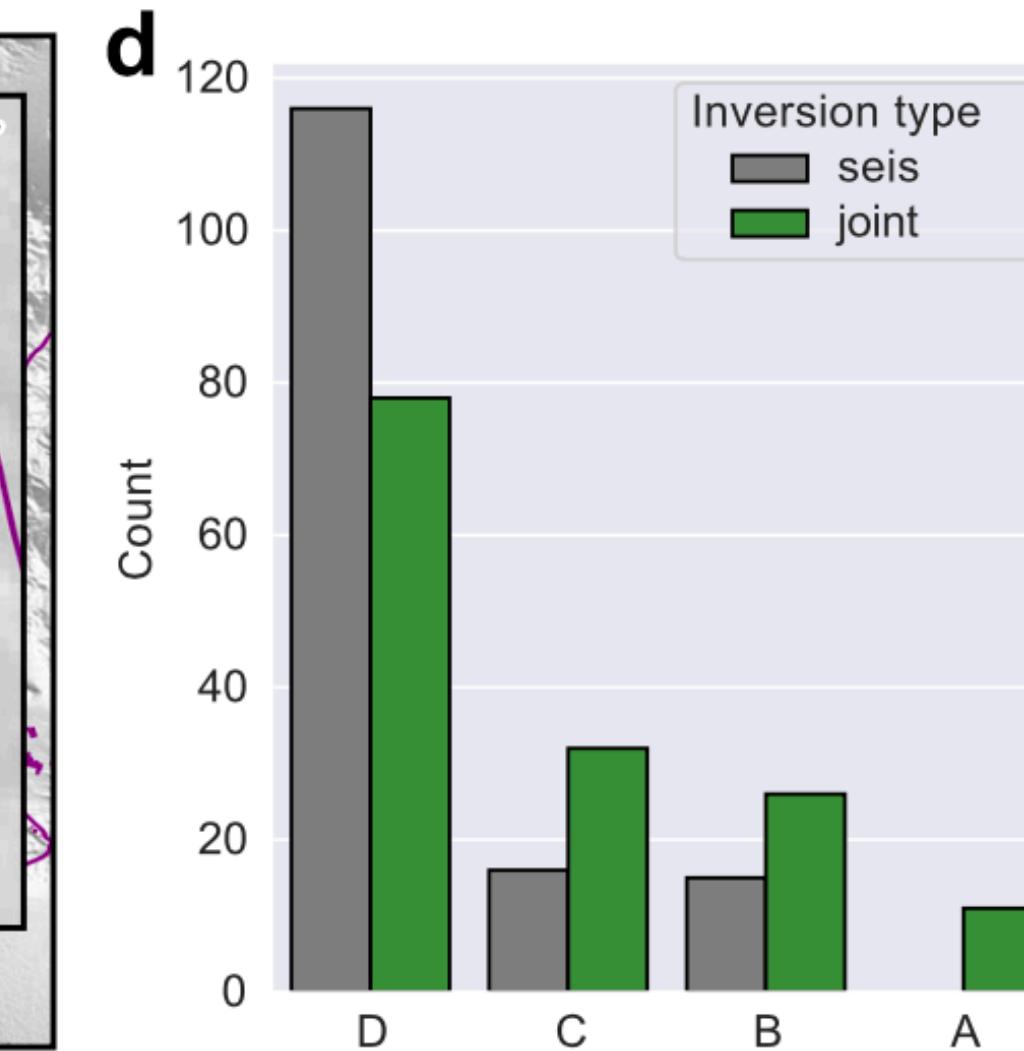
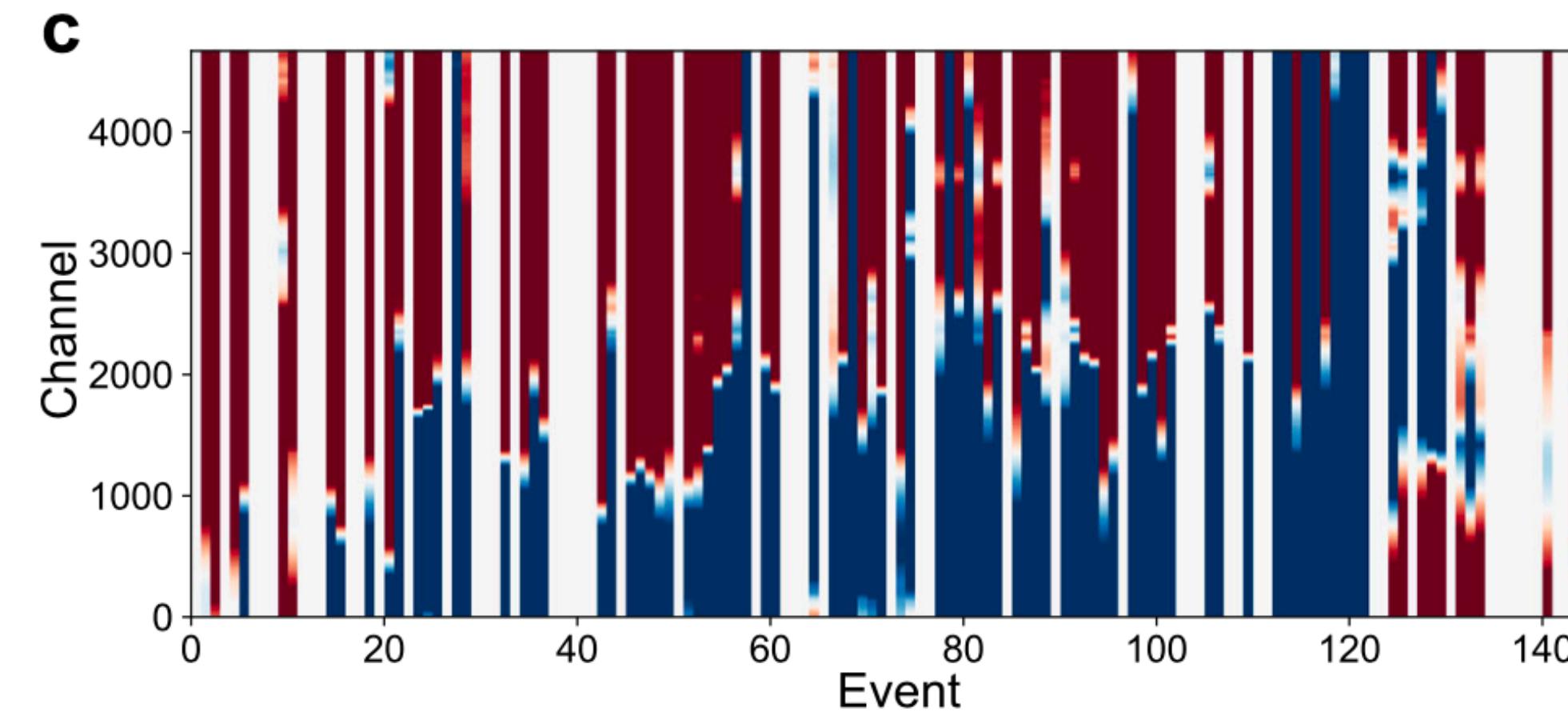
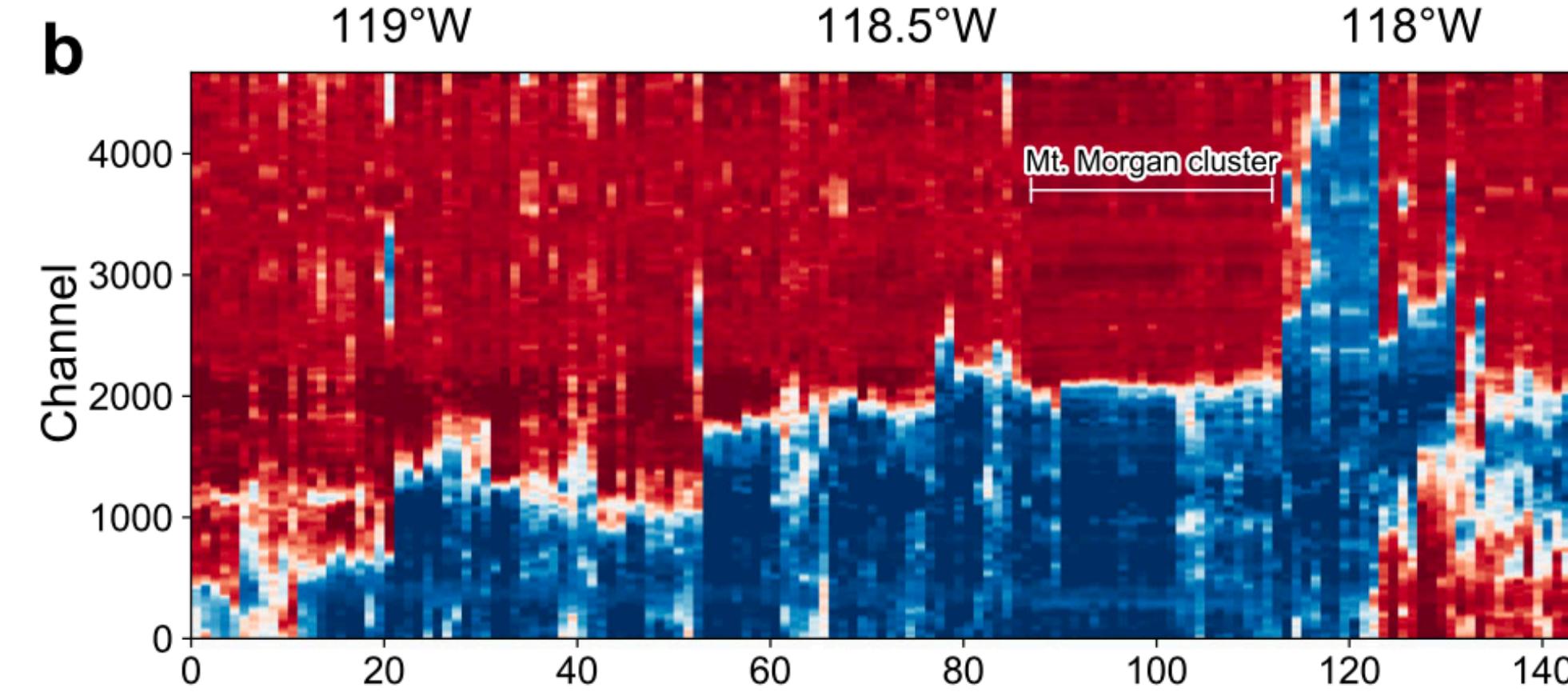
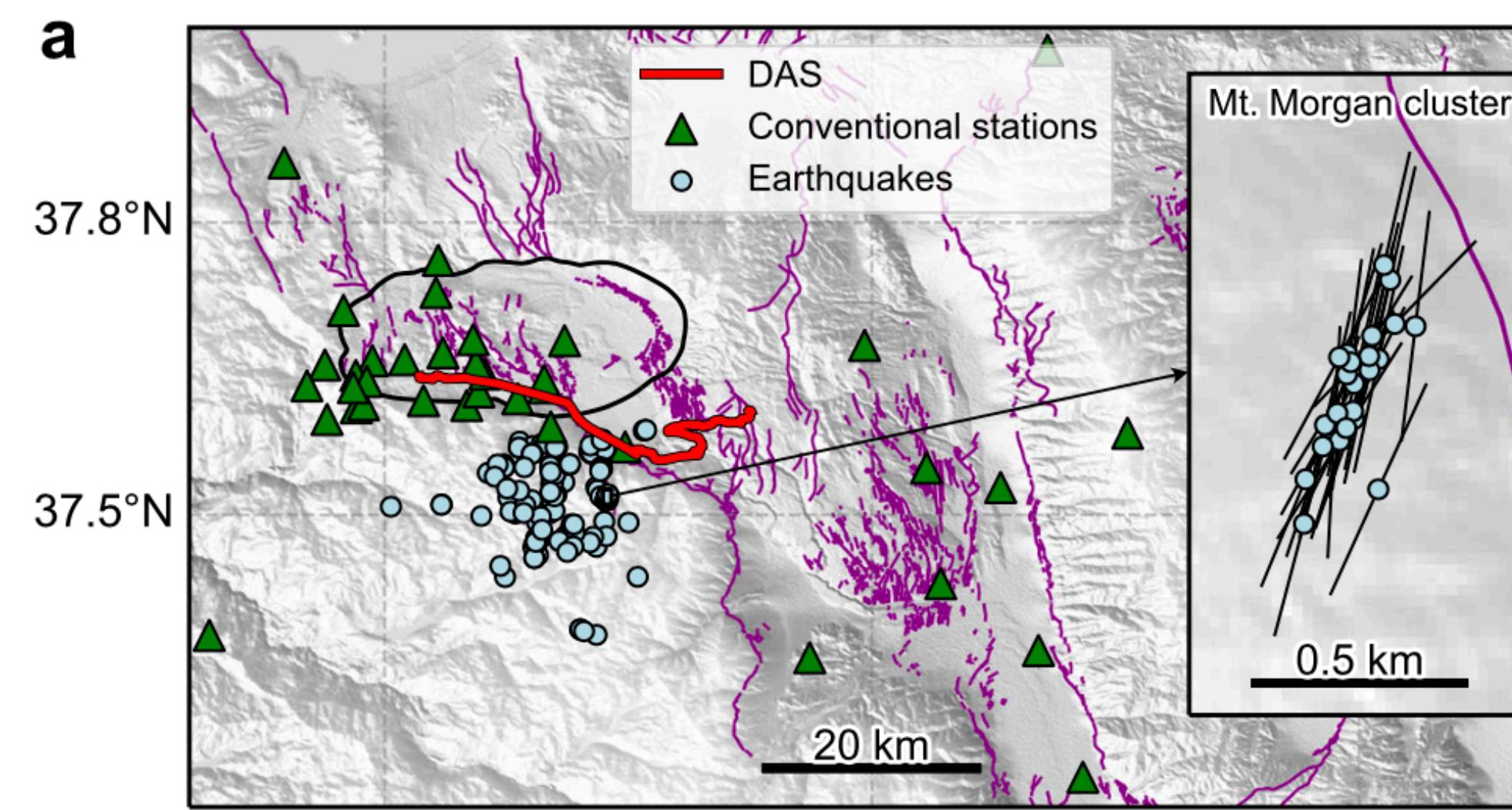


(c)

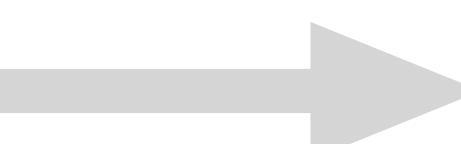


(d)

Liu and Li, 2021



# Challenges in sharing DAS data

- No standard data and metadata format
  - Data volumes are large and exceed traditional seismic data repositories
- 
- Common seismic data formats are not a good fit
  - What information is needed for metadata?

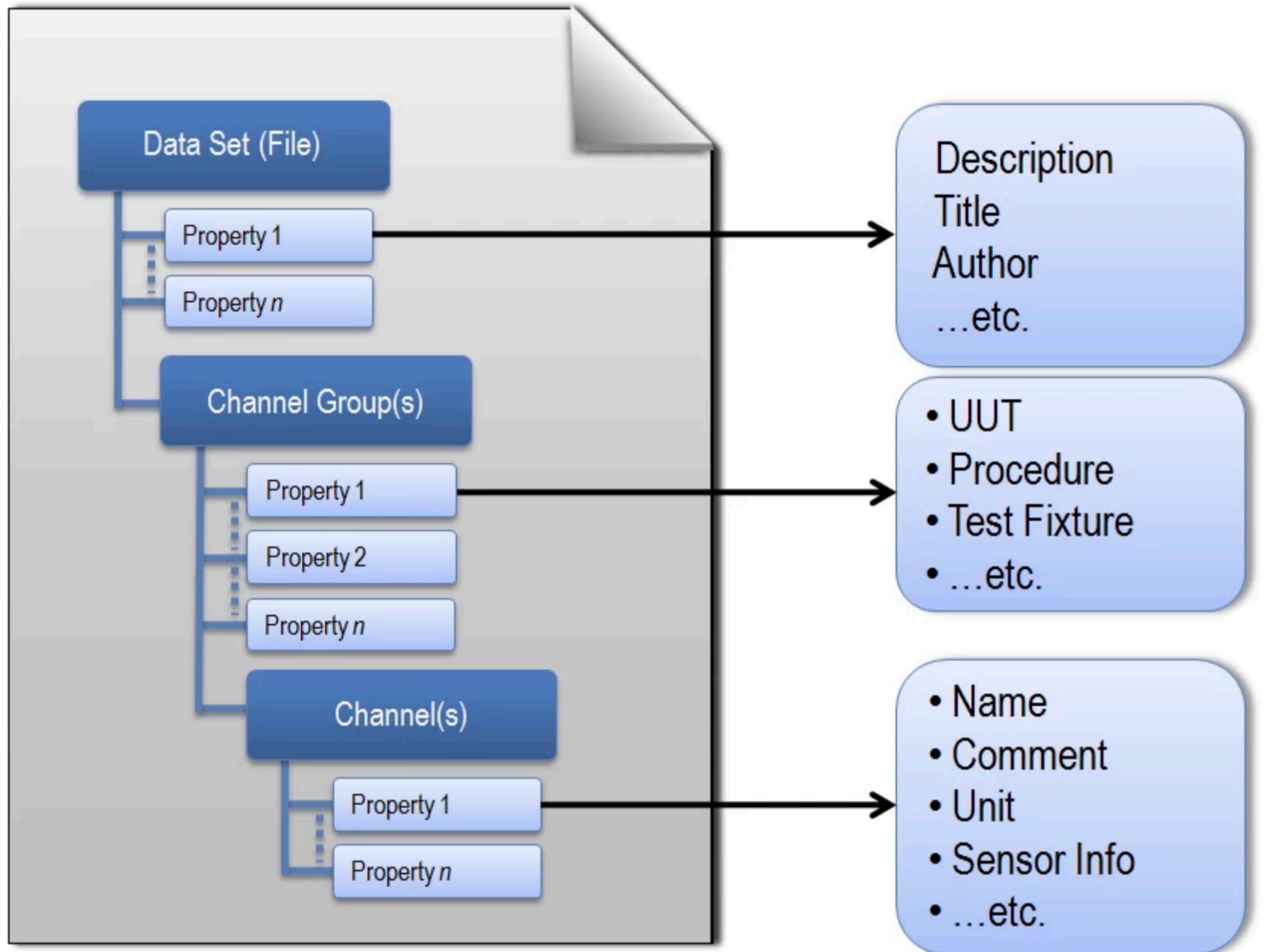
# Challenges in sharing DAS data

- No standard data and metadata format
  - Data volumes are large and exceed traditional seismic data repositories
- 
- Common seismic data formats are not a good fit
  - What information is needed for metadata?

# Data format between traditional seismic and DAS data

- Traditional seismic data:
  - SAC (Seismic Analysis Code) format.
  - MiniSEED: is the subset of the SEED that is use for time series data.
  - PH5: is a seismic data format created by IRIS PASSCAL using HDF5.
  - ASDF: The Adaptable Seismic Data Format, is a modern file format intended for researchers and analysts.
- DAS data:
  - HDF5 (OptaSense, OptoDAS, AP Sensing, etc.)
  - TDMS (Silixa iDAS)

# TDMS data structure

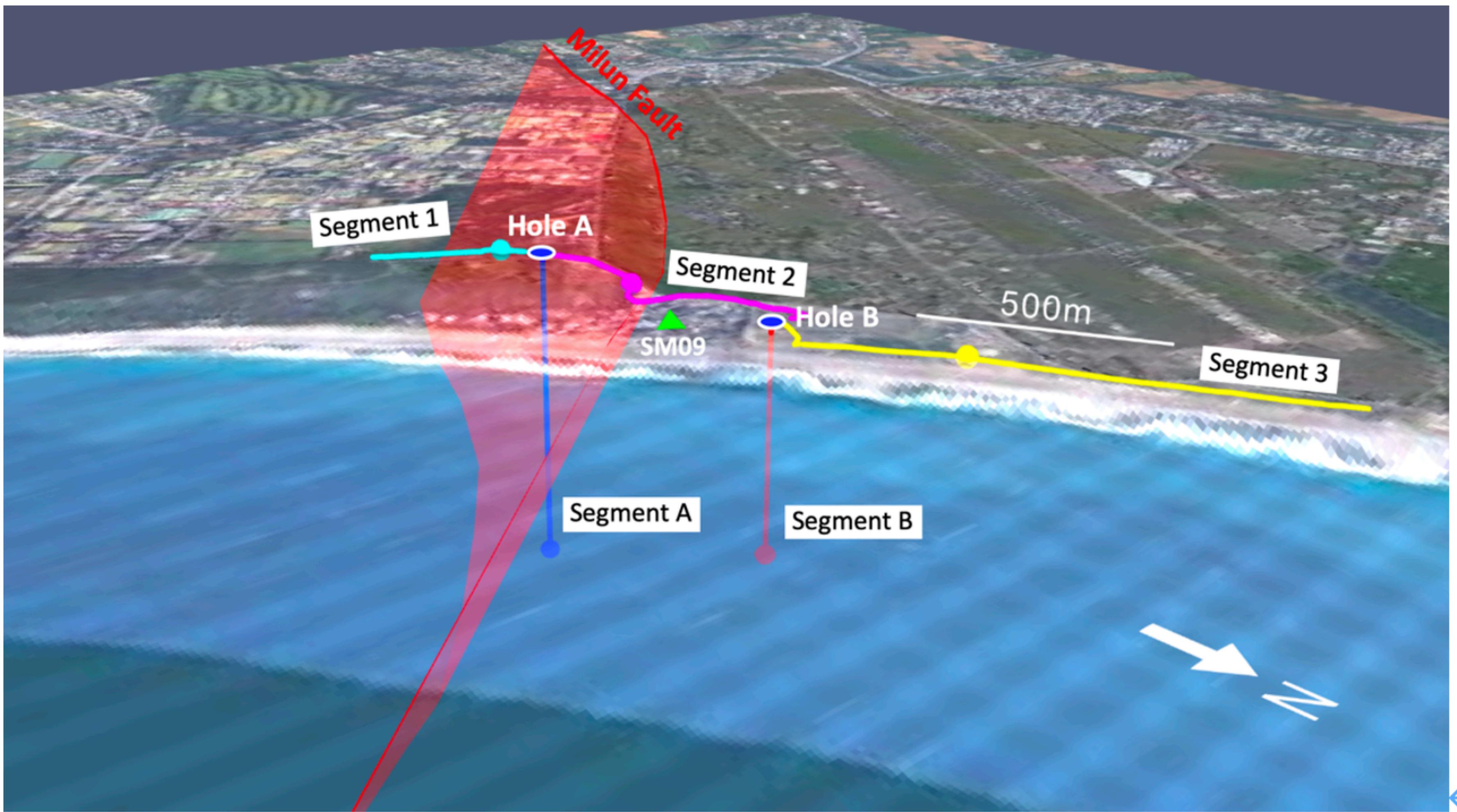


For MiDAS project:

Each TDMS file contains 20 sec waveforms records of all measurement points (nodes)

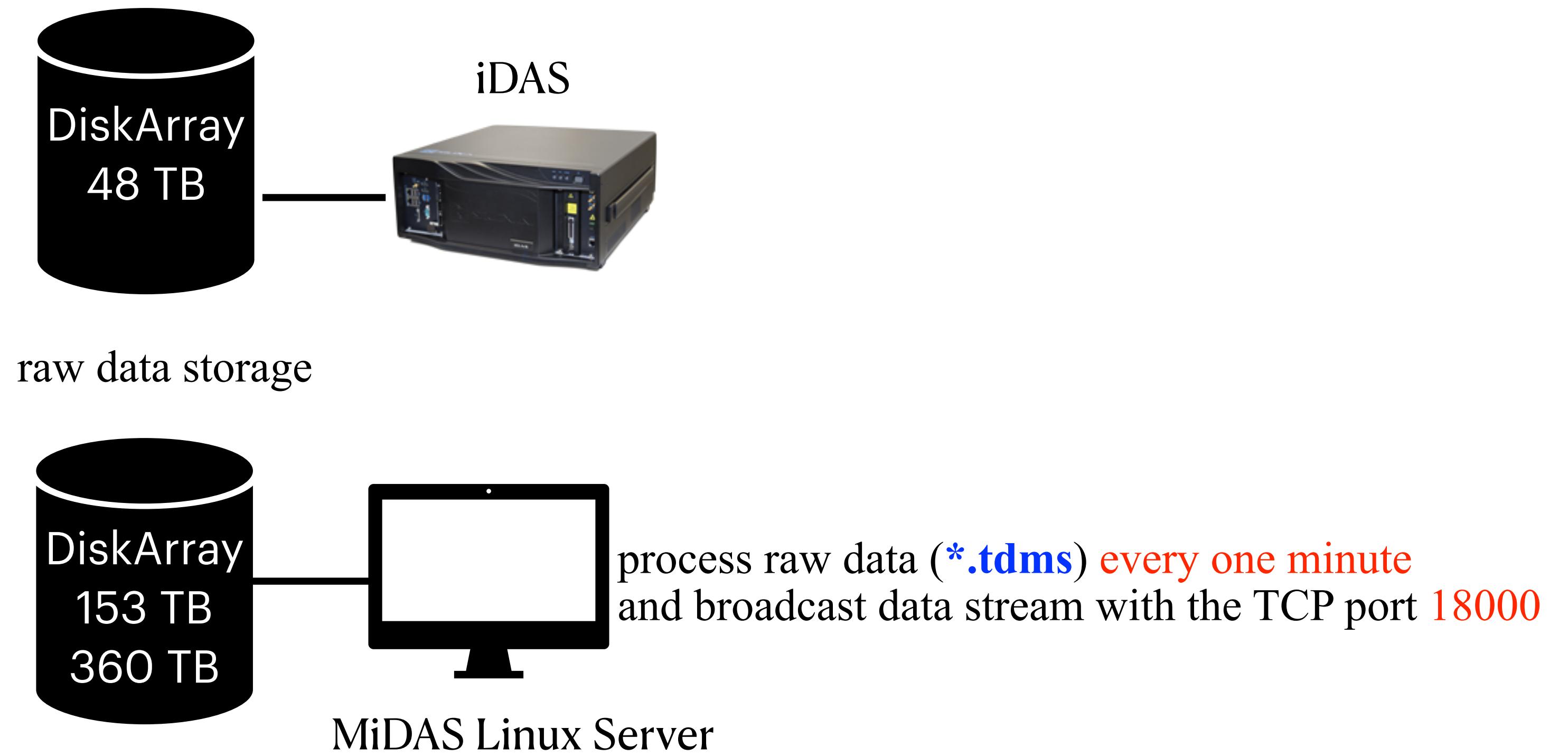
# Metadata information from raw data of iDAS

```
'name': 'MiDAS_UTC_20240102_235808.373',
'SamplingFrequency[Hz]': 1000.0,
'SpatialResolution[m]': 4.0,
'Fibre Type': 'Standard',
'Zone Type': 'Single-Zone',
'Start Distance (m)': -231.378944,
'Stop Distance (m)': 7213.405839,
'Continuous Mode': True,
'MeasureLength[m]': 7296,
'Precise Sampling Frequency (Hz)': 10000.0,
'PeakVoltage[V]': 3.3, 'Pulse 2 Delay (ns)': 500.0,
'PulseWidth[ns]': 50.0,
'Offset (m)': -231.378944, 'Receiver Gain': '70;70;70\r\n',
'Pulse Width 2 (ns)': 50.0,
'Peak Voltage 2 (V)':
'iDASVersion': '2.6.1.4', 'iDASClockSource': 'AA Onboard PLL',
'SystemInfomation.GPS.TimeReference': 'GPS',
'SystemInfomation.GPS.UTCOffset': 37.0,
'SystemInfomation.GPS.Latitude': 24.0284553,
'SystemInfomation.GPS.Longitude': 121.629038,
'SystemInfomation.GPS.Altitude': 38.719,
'SystemInfomation.GPS.SatellitesAvailable': 9,
...'
```

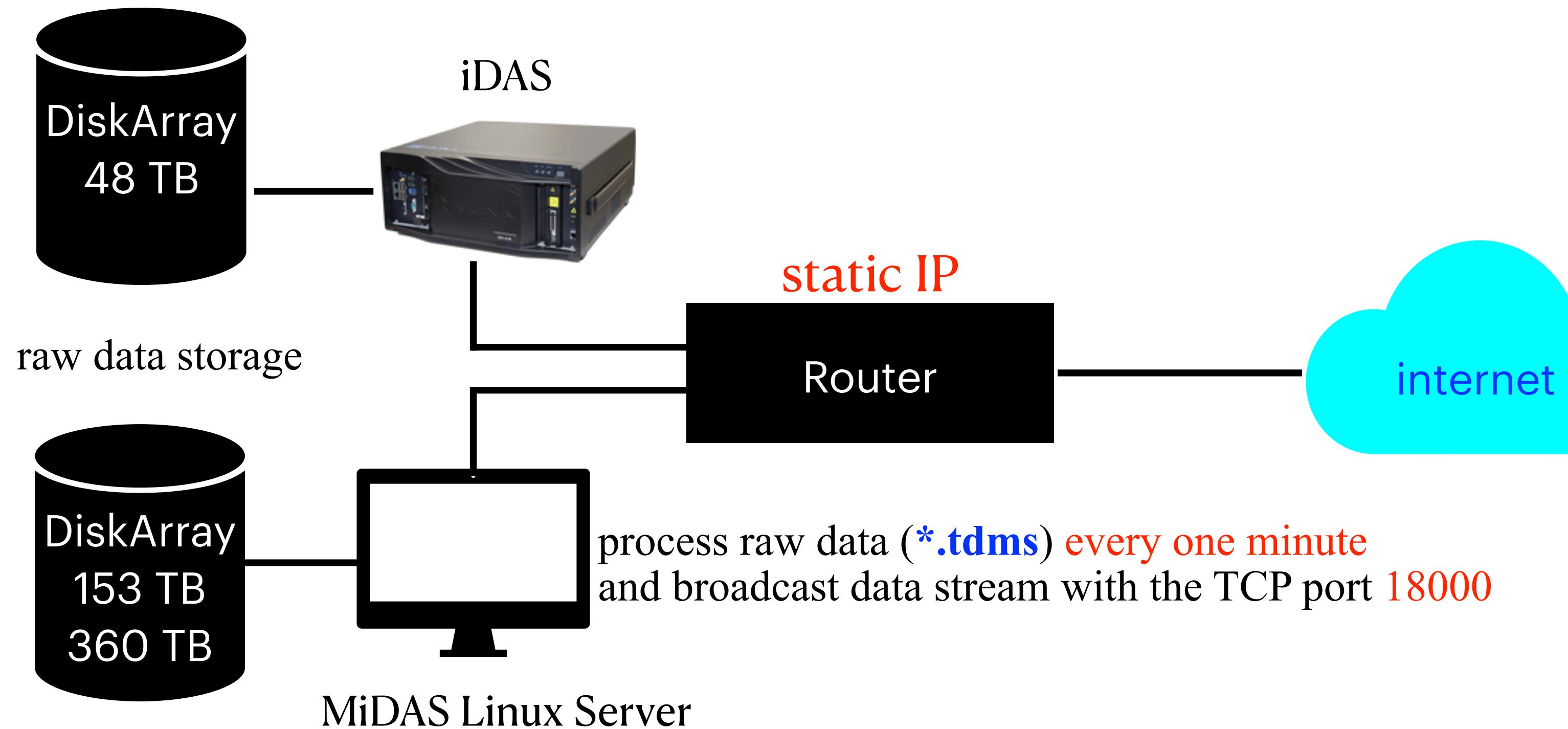


Huang et al., 2023

# Hualien



# Hualien



# Taipei

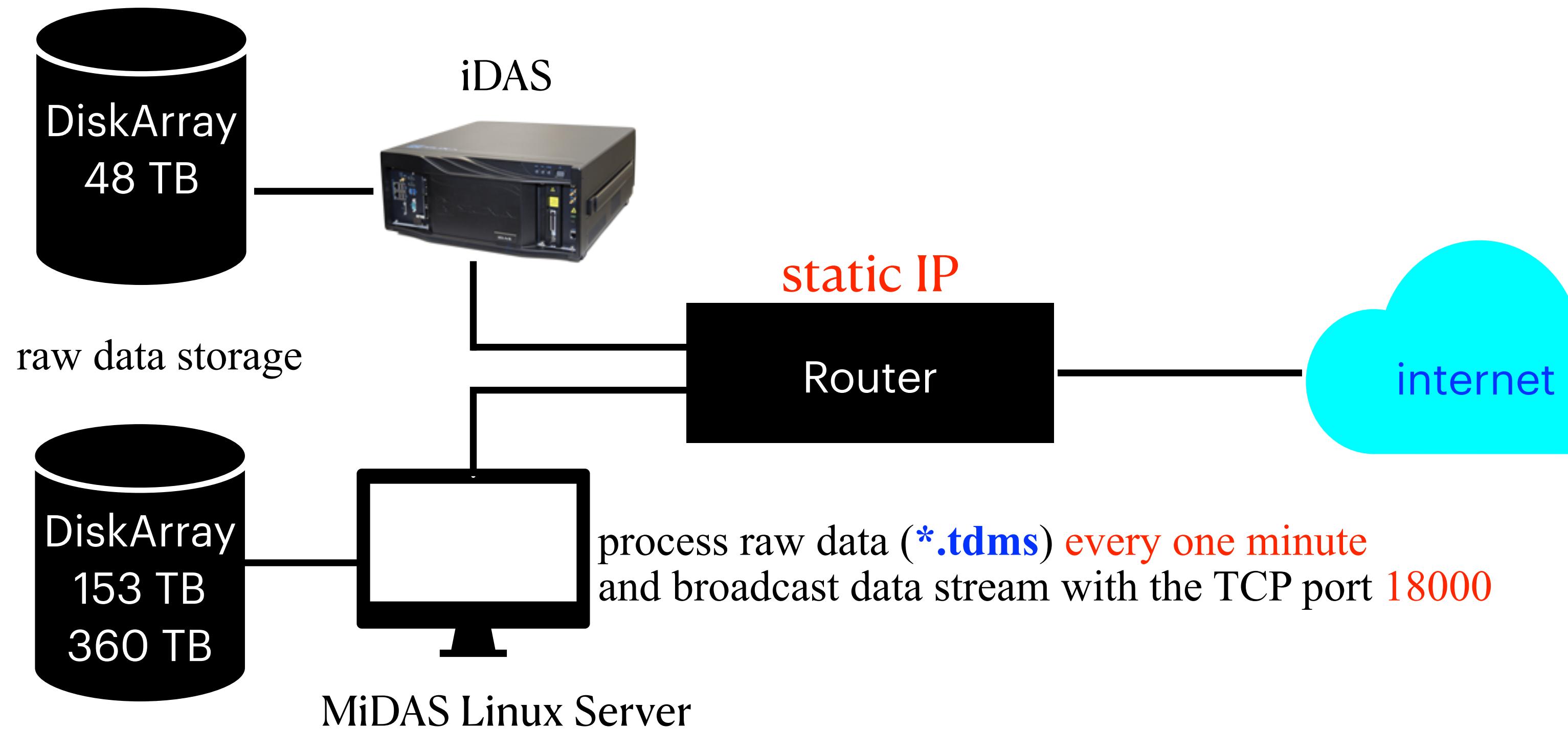
1. archive data and broadcast data stream with the TCP port **18000**
2. provide **FDSNWS** web service with port **8080**



Academia Sinica  
Grid Centre

**300 TB for hot data**

# Hualien



# Taipei

1. archive data and broadcast data stream with the TCP port **18000**
2. provide **FDSNWS** web service with port **8080**



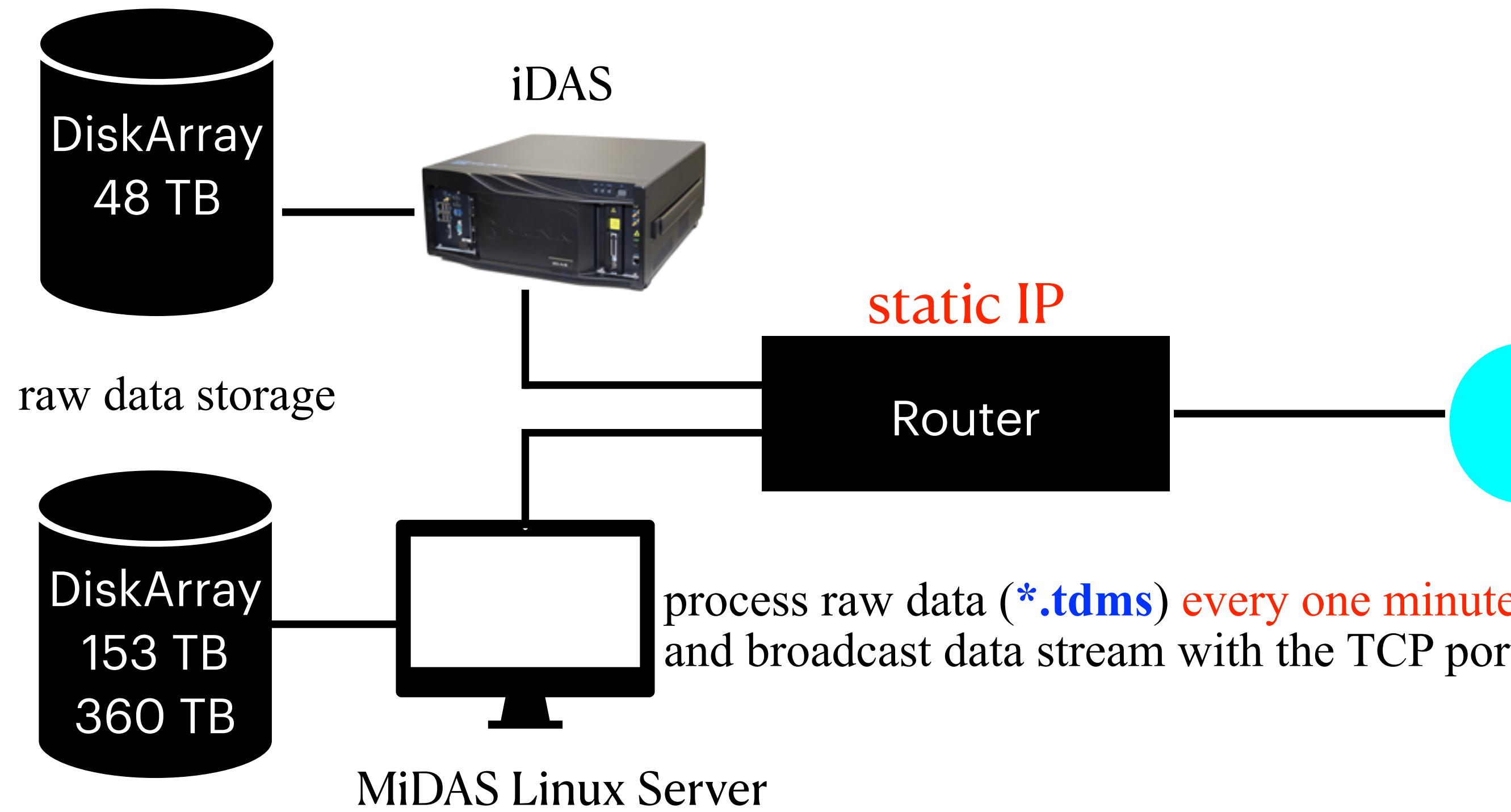
**Academia Sinica  
Grid Centre**

**300 TB for hot data**



archive data from **FDSNWS** or  
display near-real-time data

## Hualien



## Taipei

1. archive data and broadcast data stream with the TCP port **18000**
2. provide **FDSNWS** web service with port **8080**



Academia Sinica  
Grid Centre

**300 TB for hot data**

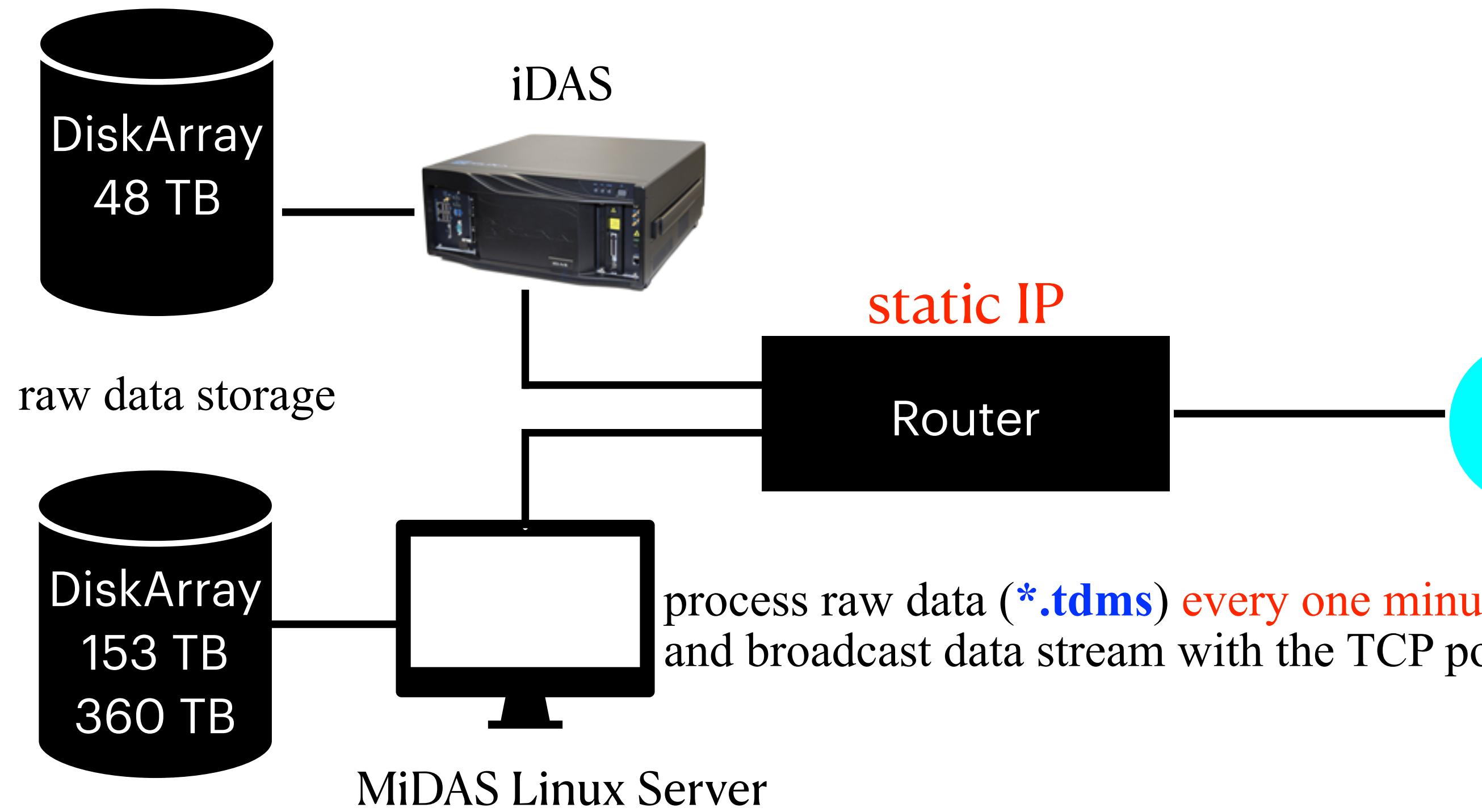


for public



archive data from **FDSNWS** or  
display near-real-time data

# Hualien



## Data capacity

1824 stations (~7.3 km fiber)  
~300 GB/day = ~8.8TB/month

# Taipei

1. archive data and broadcast data stream with the TCP port 18000
2. provide FDSNWS web service with port 8080



Academia  
Sinica  
Grid  
Centre

**300 TB for hot data**

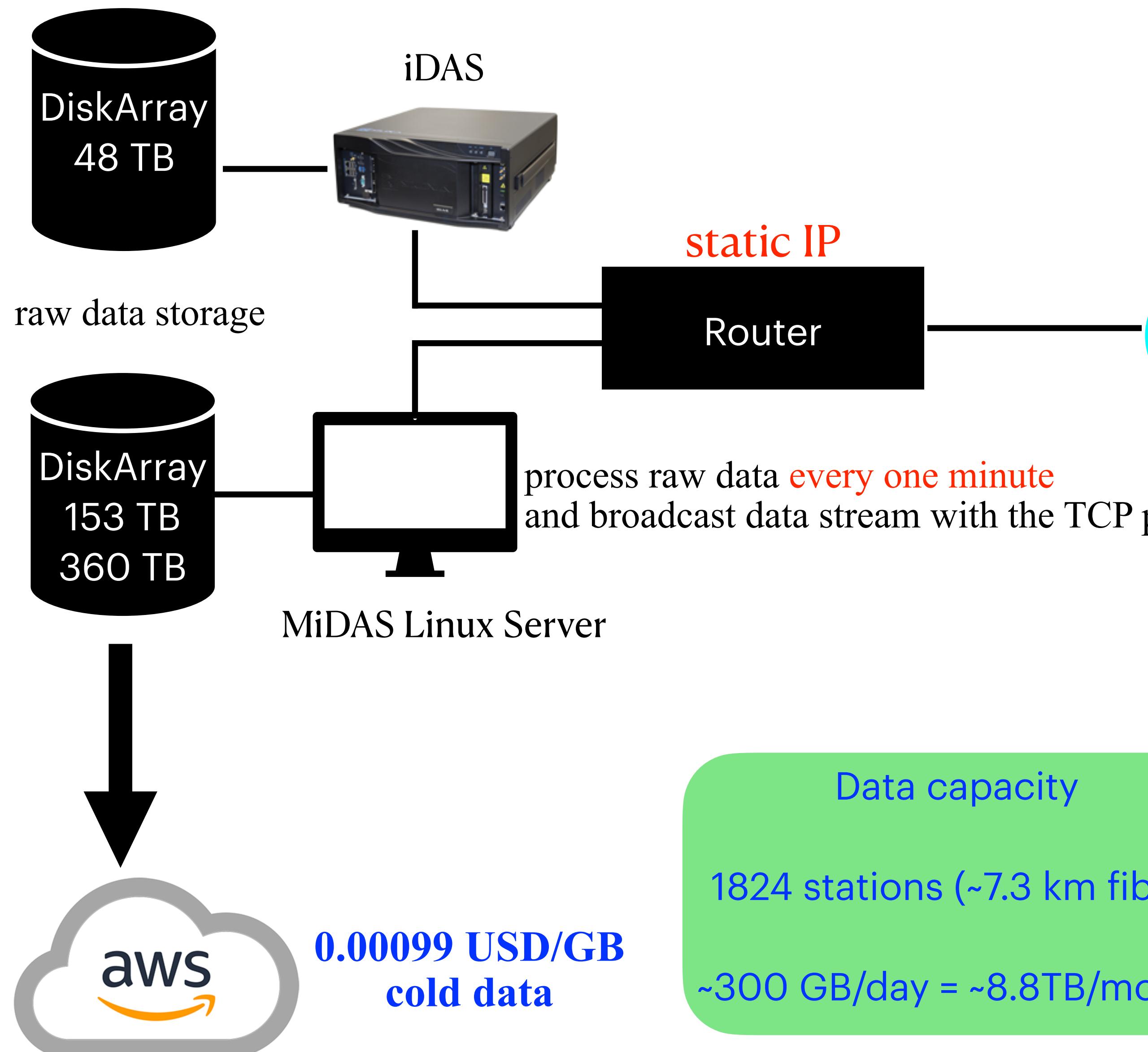


for public



archive data from FDSNWS or  
display near-real-time data

# Hualien



# Taipei

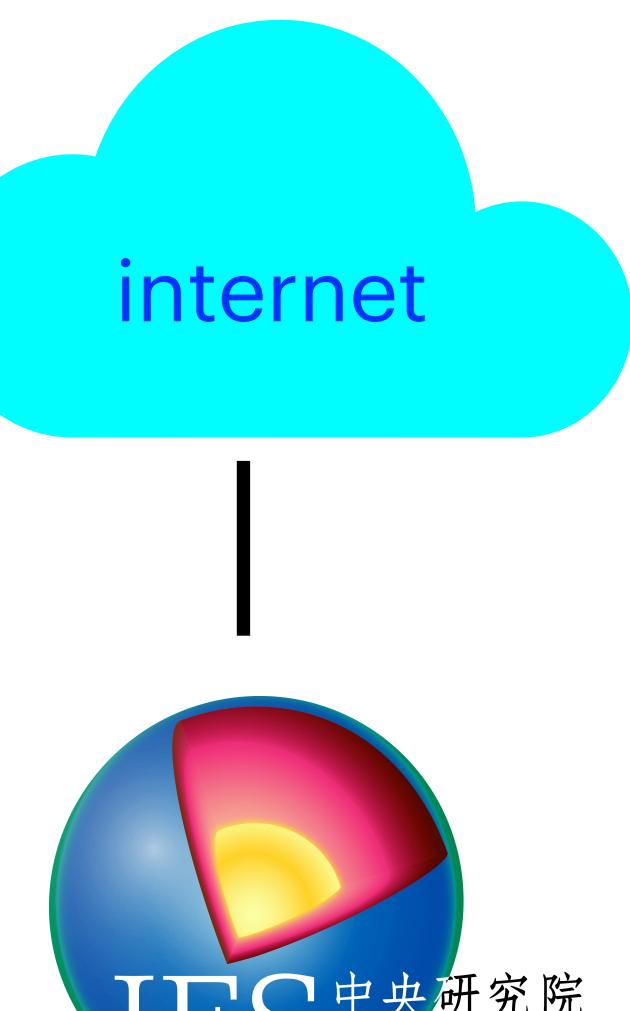
1. archive data and broadcast data stream with the TCP port **18000**
2. provide **FDSNWS** web service with port **8080**



**300 TB for hot data**



for public



archive data from **FDSNWS** or  
display near-real-time data

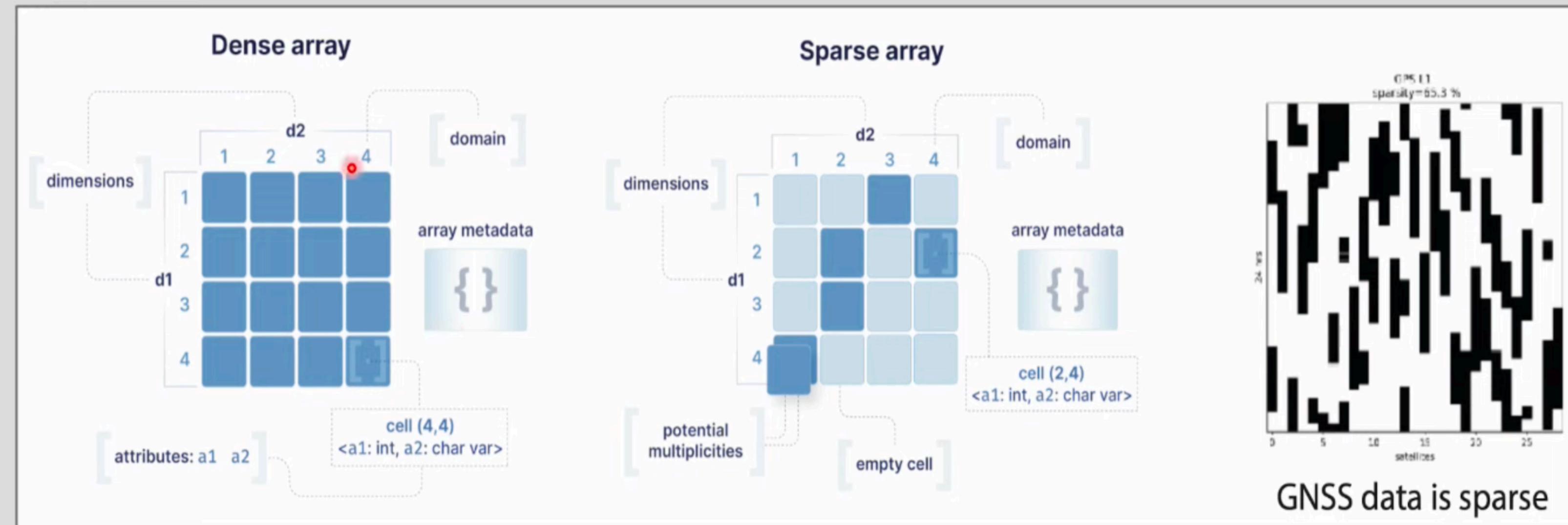
# SeisComP Data Structure (SDS)

- <SDSdir>/Year/NET/STA/CHAN.TYPE/NET.STA.LOC.CHAN.TYPE.YEAR.DAY
- archive/2023/TW/00000
  - 00001
  - 00002
  - ~
  - 01822
  - 01823/HSF.D/TW.01823..HSF.D.2024.001
    - TW.01823..HSF.D.2024.002
    - ~
    - TW.01823..HSF.D.2024.011
    - TW.01823..HSF.D.2024.012

# Data Container (TileDB)



- Designed for cloud (object store) storage systems from first principles (open specification/source)
- Performant multidimensional slicing via tiling
- Support for sparse arrays
- Supports versioning (by time)
- Multi-threaded Read/Write directly to disk or object storage (S3, ...)

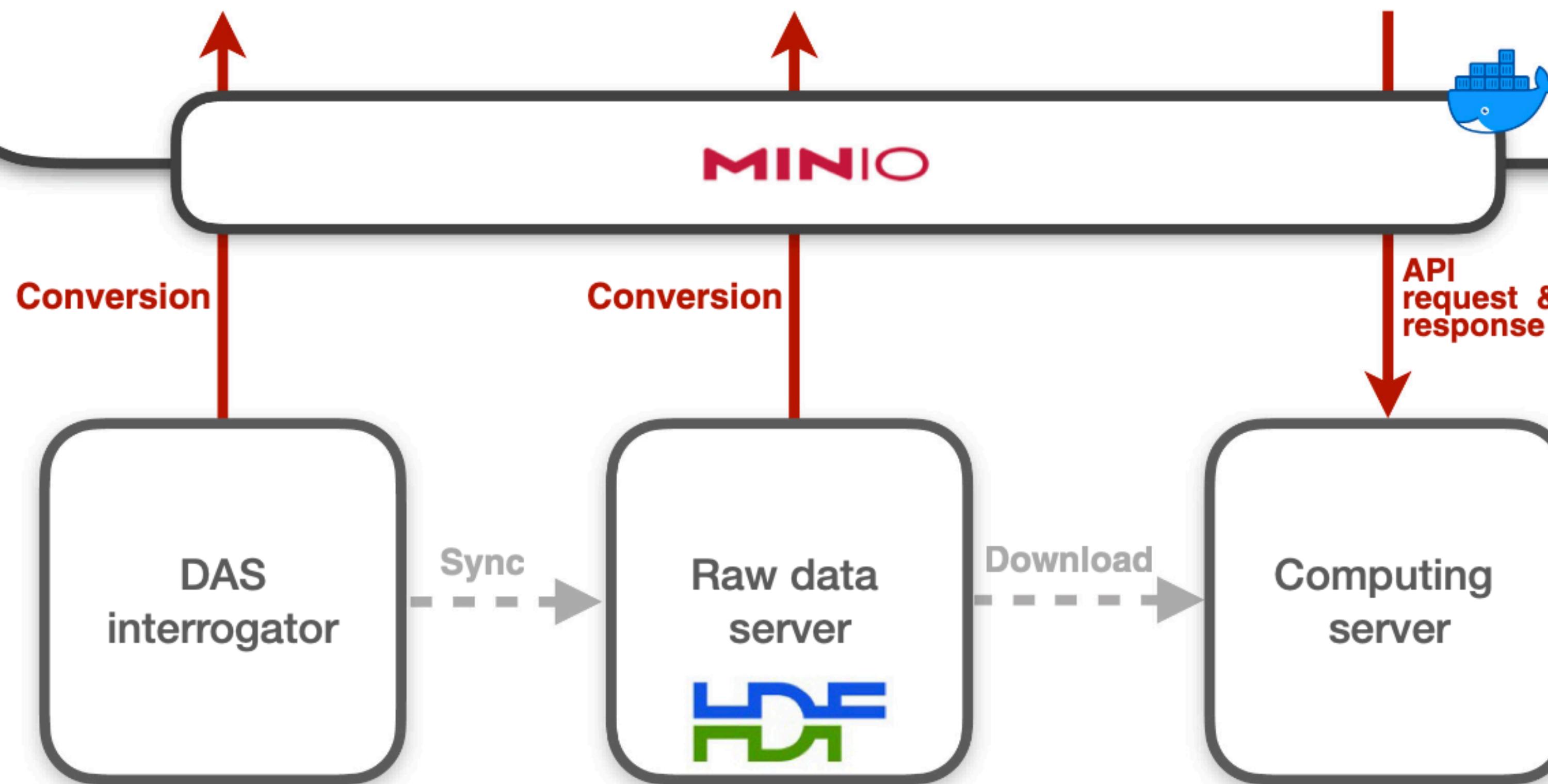
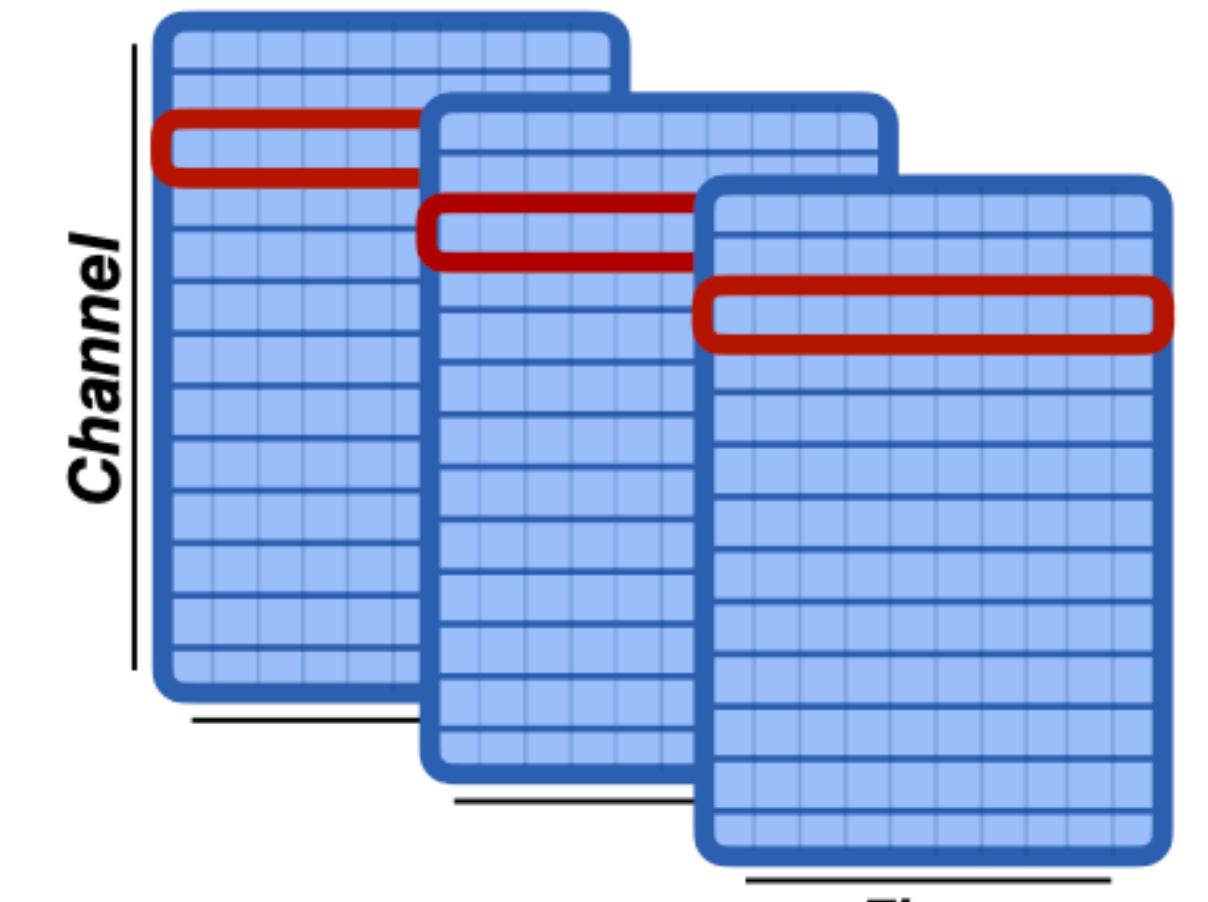


TileDB is a library that efficiently manages large-scale, n-dimensional, dense and sparse array data.

# Object storage server

[tile]DB  
-----  
Zarr

- RawData
- RawDataTime
- GpBits
- GpsStatus
- PpsOffset
- SampleCount



Ni et al., 2023



globus

a uchicago non-profit service



GET STARTED



LOG IN

Solutions ▾

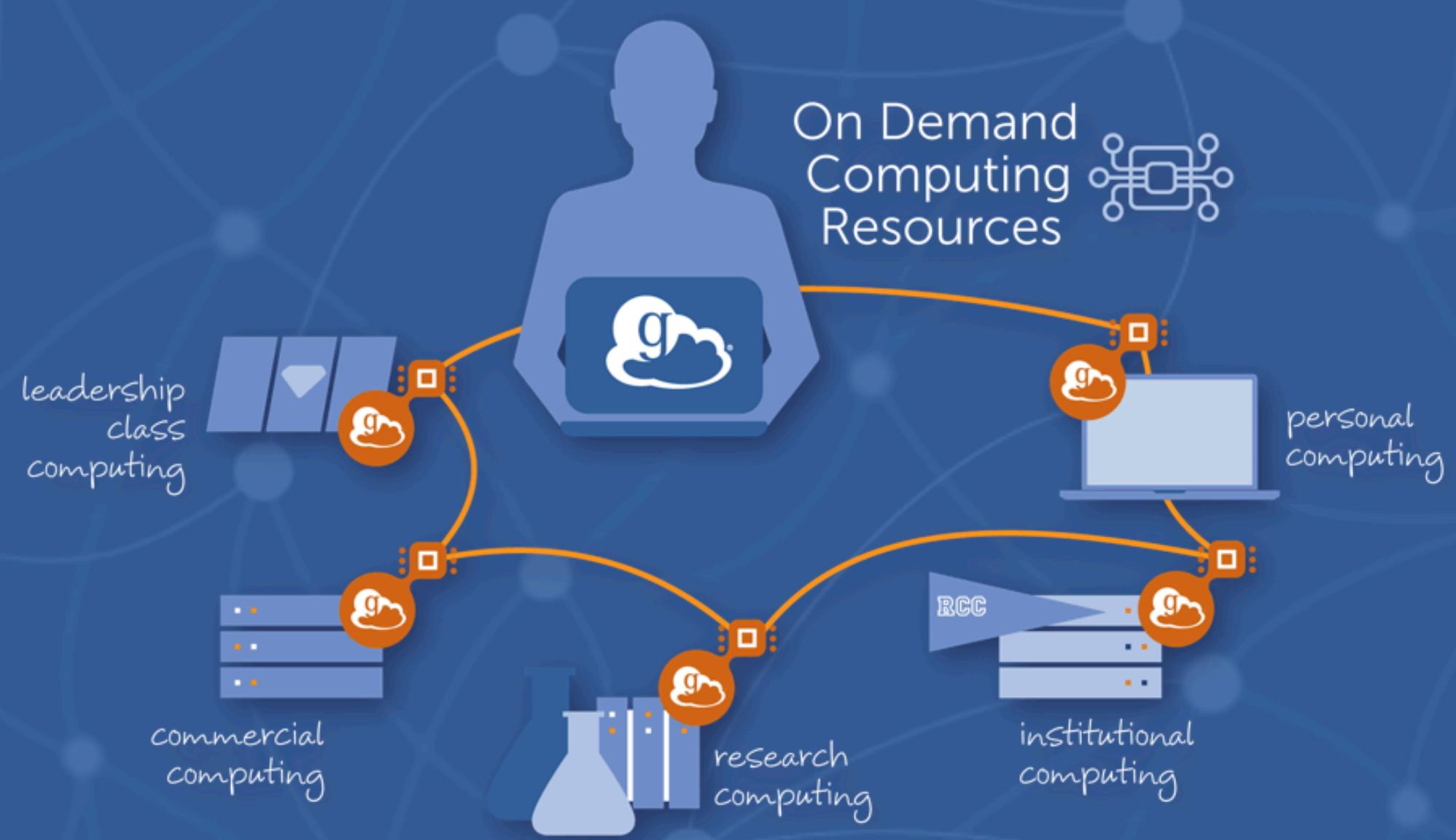
Resources ▾

Pricing ▾

Newsroom ▾

Developers ▾

About ▾

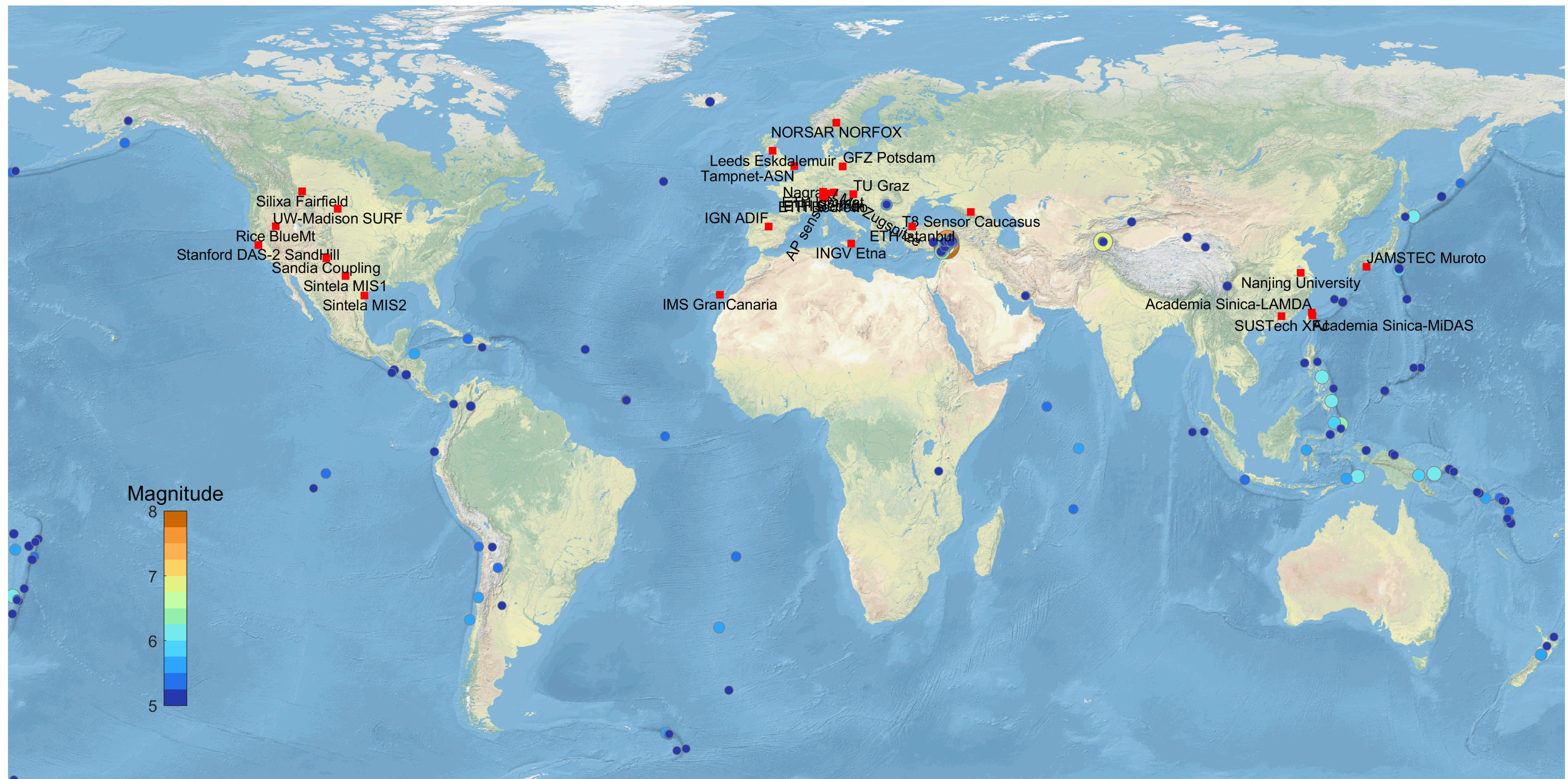


Go beyond data

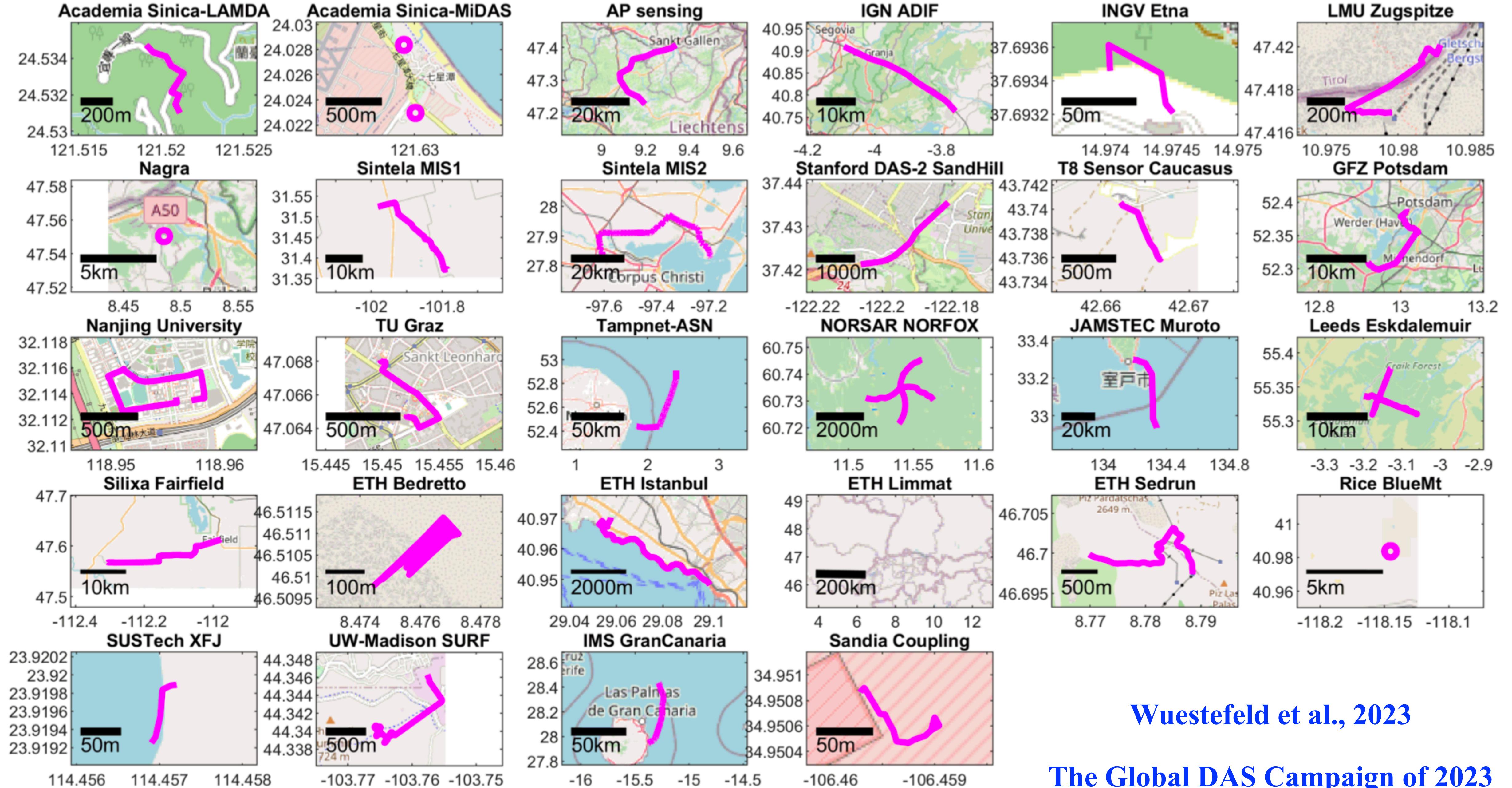
# Globus Compute

Reliable, distributed Function-as-a-Service

COMPUTE ANYWHERE: EDGE TO SUPERCOMPUTER ➔



The Global DAS Campaign of 2023 (Wuestefeld et al., 2023)



Wuestefeld et al., 2023

The Global DAS Campaign of 2023

# Summary

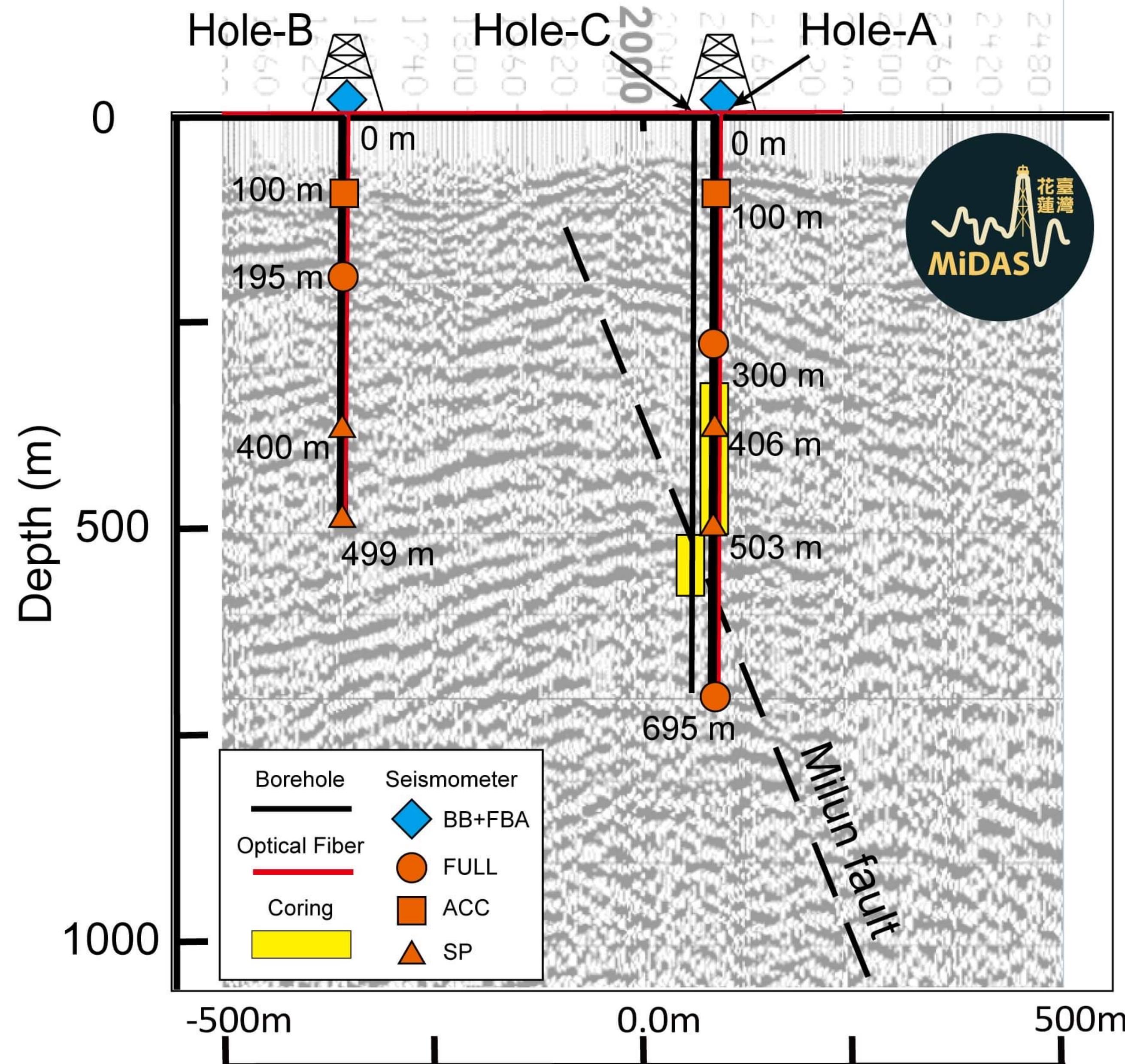
- High resolution of DAS data provides the new era of the seismic observation
- The near-real-time DAS database has been constructed for MiDAS project
- Effectively sharing DAS data remains the most critical issue at present
- The Cloud application (including computing and storage) is an inevitable step for DAS data processing

A photograph of a serene sunset over a large body of water. The sky is a gradient from deep blue at the top to warm orange and yellow near the horizon. In the foreground, several small boats are scattered across the dark water. On the right side, the silhouette of a large tree branch is visible against the sky. In the distance, a range of mountains is faintly visible.

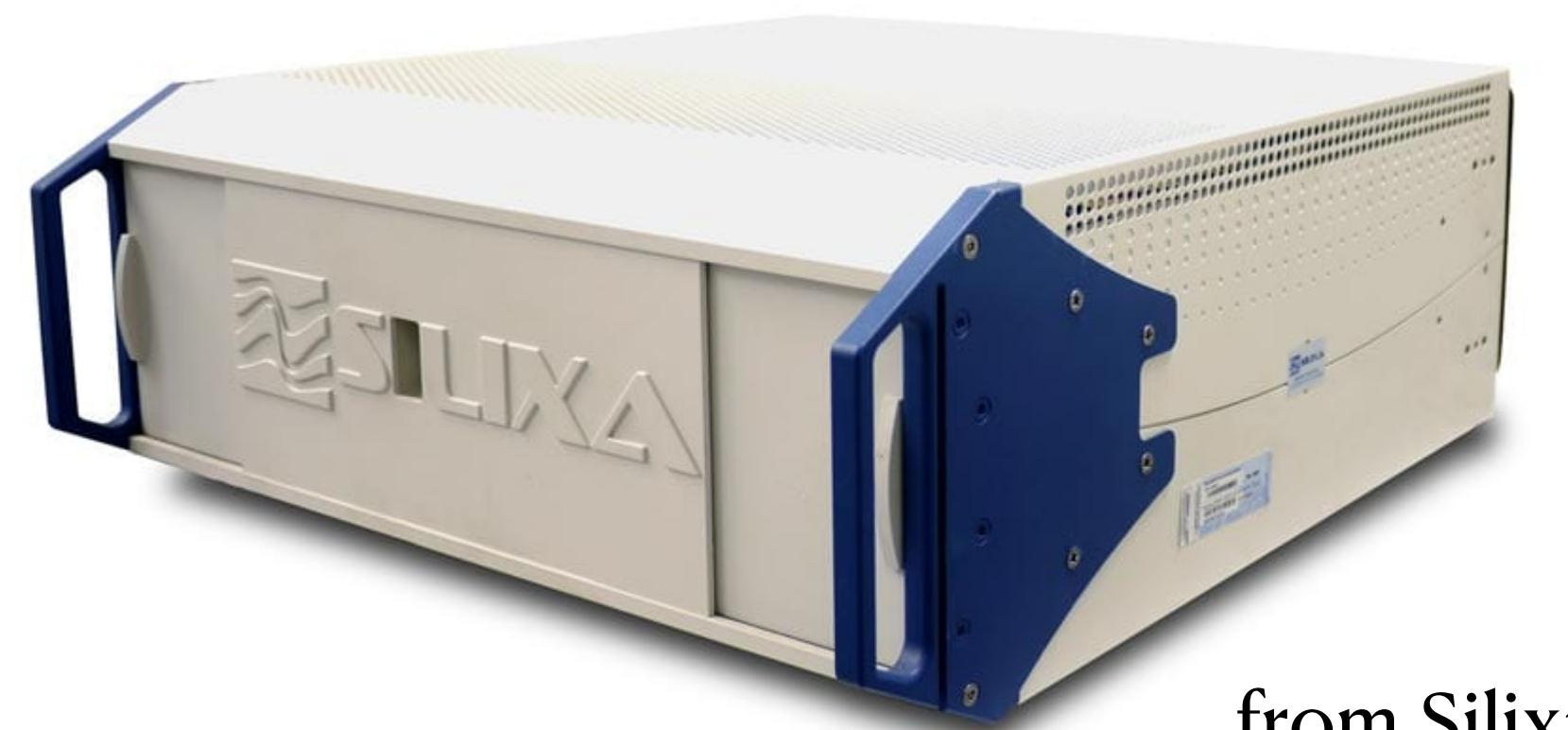
**Thank you for your attention!**



Courtesy to Chin-Jen Lin



Courtesy to Yen-Yu Lin



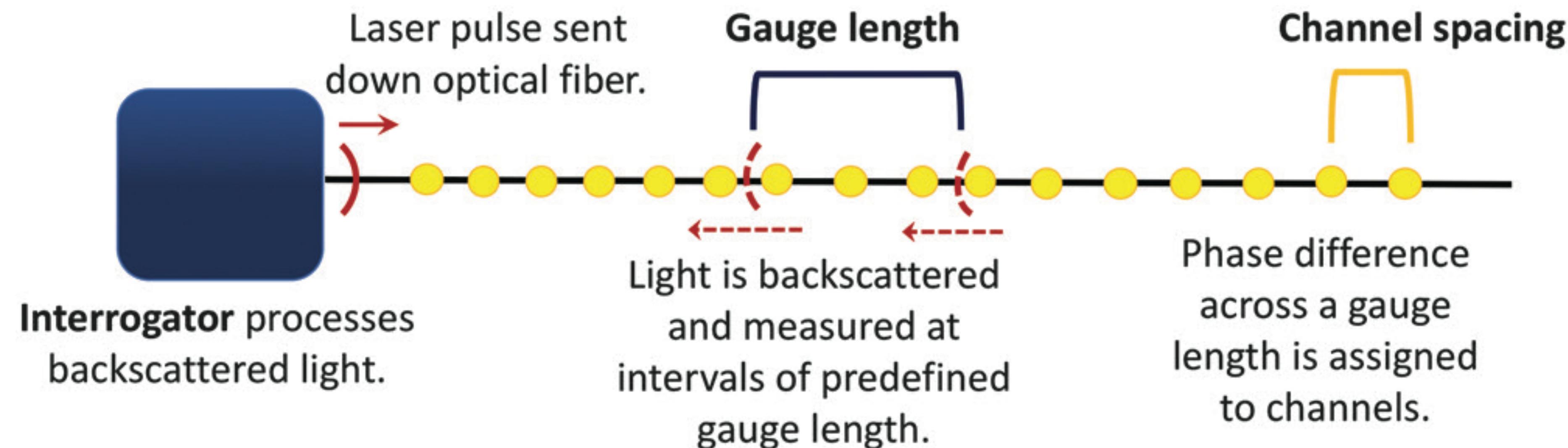
from Silixa Ltd



Fiber of Hole A was deployed on 2021/12/31



Courtesy to Chin-Jen Lin



Courtesy to APsensing

Gauge length: **spatial resolution**, which is generally larger than the sampling resolution  
 Channel spacing: the distance between consecutive measurement points (**sampling resolution**)

for MiDAS project: gauge length=10m; channel spacing=4m ; smapling rate=1000Hz  
**1824 stations (~7.3 km optical fiber)**