Data Committee Update
8 January, 2024

Tim Ahern, IRIS/Earthscope Emeritus, ESO Data Committee

Joel Simon, Earthscope-Oceans

Cristoph Waldman, MARUM

EARTHSCOPE-OCEANS DATA COMMITTEE

- Joel Simon, Princeton
  - Now Chair of FDSN WG V
  - Responsible for the effort to get MH data into GeoCSV format
    - GeoCSV files for MH can be found at http://ds.iris.edu/data/reports/MH/
- Christoph Waldman, Marum
- We would like to add a member from SUSTECH.
  - Yong Yu will be invited to join the committee

SHIPMENT OF US AND FRENCH MERMAID DATA IN 2023

- 1.69 gigabytes shipped to 16 distinct locations

- Shipped to 5 countries
  - US 75.5%   France 20.8%   China 3.5%   Hong Kong 0.3%   Japan <0.1%
  - 99.7% shipped FDSN Dataselect web service

station	Bytes	Gibibytes	Percent
P0008	402,280,448	0.374653 GiB	22.2 %
P0010	362,942,464	0.338017 GiB	20.0 %
P0006	262,496,256	0.244469 GiB	14.5 %
P0009	151,650,304	0.141235 GiB	8.4 %
P0012	98,885,632	0.092094 GiB	5.5 %
P0011	97,095,680	0.090427 GiB	5.4 %
P0013	86,675,456	0.080723 GiB	4.8 %
P0023	69,312,512	0.064552 GiB	3.8 %
P0016	58,474,496	0.054459 GiB	3.2 %
P0020	32,935,936	0.030674 GiB	1.8 %
P0022	32,677,888	0.030434 GiB	1.8 %
P0007	31,440,896	0.029282 GiB	1.7 %
P0017	26,902,528	0.025055 GiB	1.5 %
P0021	25,358,336	0.023617 GiB	1.4 %
P0024	21,266,432	0.019806 GiB	1.2 %
P0019	19,259,392	0.017937 GiB	1.1 %
P0018	18,132,992	0.016888 GiB	1.0 %
P0025	13,217,792	0.012310 GiB	0.7 %

DATA COMMITTEE TOPICS

- GeoCSV update

- Proposed 2024 ESO Data Policy

IDENTIFIED GEOCSV COMPONENTS

- GeoCSV Header

- Column Identifiers

- A series of Elements as needed
  - Time and SNCL Information (normally necessary)
    - Start time, network, Station, Location, Channel
  - Positional Information
    - Latitude, Longitude, Elevation, Depth
  - Sensor information
    - SensorDescription, Scale, Scale Frequency, Scale Units
  - Timing Information
    - SampleRate, TimeDelay, TimeCorrection
  - Orientation Information
    - Dip, Azimuth, Uncertainties

A GEOCSV EXAMPLE FROM P0008

GEOCSV HEADER

GeoCSV Header					
Column Row Headers					
Time + SNCL	Positional	Sensor/Gain	Timing	Orientation	As Needed

- GeoCSV Header
  - #dataset: GeoCSV 2.0
  - #created: 2021-07-28T23:25:20Z
  - #automaid: v3.4.2 (https://github.com/earthscopeoceans/automaid (doi: 10.5281/zenodo.5057096))
  - #delimiter: ','
  - #lineterminator: '\n'

COLUMN ROW HEADERS

GeoCSV Header					
Column Row Headers					
Time + SNCL	Positional	Sensor/Gain	Timing	Orientation	As Needed

degrees_north	degrees_east	meters	meters
float	float	float	float
latitude	longitude	elevation	depth

POSSIBLE ELEMENTS – ESO EXAMPLE

Sensor orientation for MH OBS could go here

Header Column Row Headers

Type	Time and SNCL				Position				Sensor				Timing		
#field_unit	ISO_8601	unitless	unitless	unitless	degrees_nort	degrees_east	meters	meters	unitless	factor	hertz	unitless	hertz	seconds	seconds
#field_type	datetime	string	string	string	float	float	float	float	string	float	float	string	float	float	float
MethodIden	StartTime	Network	Station	Location	Channel	Latitude	Longitude	Elevation	Depth	SensorDescri	Scale	ScaleFreque	ScaleUnits	SampleRate	TimeDelay	TimeCorrection

Data

Measuremen	2018-08-05T	MH	P0008		nan	-12.008233	-172.0231	0	0	MERMAIDHy	nan	nan		nan	0.00003	nan
Measuremen	2018-08-05T	MH	P0008		nan	-12.006967	-172.01872	0	0	MERMAIDHy	nan	nan		nan	-0.000062	nan
Measuremen	2018-08-06T	MH	P0008		nan	-12.0477	-172.01357	0	0	MERMAIDHy	nan	nan		nan	0.437377	nan
Measuremen	2018-08-06T	MH	P0008		nan	-12.047684	-172.01369	0	0	MERMAIDHy	nan	nan		nan	0	nan
Measuremen	2018-08-06T	MH	P0008		nan	-12.047584	-172.01425	0	0	MERMAIDHy	nan	nan		nan	0	nan
Algorithm:au	2018-08-08T	MH	P0008	0	BDH	-12.074427	-171.99651	0	1531	MERMAIDHy	-149400	1	Pa	20	nan	-0.29087
Measuremen	2018-08-15T	MH	P0008		nan	-12.205566	-171.90376	0	0	MERMAIDHy	nan	nan		nan	1.643707	nan
Measuremen	2018-08-15T	MH	P0008		nan	-12.2059	-171.90405	0	0	MERMAIDHy	nan	nan		nan	0	nan
Measuremen	2018-08-15T	MH	P0008		nan	-12.207367	-171.9055	0	0	MERMAIDHy	nan	nan		nan	-0.000184	nan
Measuremen	2018-08-15T	MH	P0008		nan	-12.207784	-171.90599	0	0	MERMAIDHy	nan	nan		nan	0	nan
Algorithm:au	2018-08-16T	MH	P0008	0	BDH	-12.231257	-171.89015	0	1527	MERMAIDHy	-149400	1	Pa	20	nan	-0.242693
Algorithm:au	2018-08-17T	MH	P0008	0	BDH	-12.255579	-171.86842	0	1521	MERMAIDHy	-149400	1	Pa	20	nan	-0.523292
Measuremen	2018-08-17T	MH	P0008		nan	-12.2612	-171.86581	0	0	MERMAIDHy	nan	nan		nan	0.569366	nan

CURRENT STATUS

- 2023 FDSN meetings at IUGG in Berlin
  - GeoCSV officially adopted as an FDSN Standard
  - In order to maximize use the goal is to merge the GeoCSV information into the existing StationXML standard
  - NSF has been contacted
    - Submit a short description of the project
    - Potential to fund small proposal outside of normal solicitation process and deadlines
    - My current estimate is about $100K (not including overheads) to
      - Develop an application that will insert GeoCSV information files directly into StationXML
      - Extract information from StationXML and output compliant GeoCSV files in a lossless manner

NEXT STEPS FOR GEOCSV

- Find funding (NSF has been contacted)
- Two potential contractors currently developing cost proposals
- Select contractor
- Select host organization to support the proposal
  - Ronin Institute (low overhead) but need to become Ronin Scholar
  - EarthScope or Princeton (higher overhead but less difficult)
  - Another not-for-profit
- Write the proposal (a rough outline has been developed)
- Proposal will support application development and travel support to coordinate proposal adoption and outreach at targeted international and national meetings (IASPEI, AGU, etc.)
- Target release by August 2025 IASPEI/FDSN meeting

DATA POLICY REDO

- 2021 Proposal

- **Recommended Data Policy and Citation**

- The data policy established by ESO is quite progressive. Data from at least 10% of an ESO partners MERMAIDS should be released without any delay other than the time required for data curation.

- Data from all other stations must be released after two years.

- Reference ESO data using DOI 10.7914/SN/MH.

But some members had concerns with the above so…..

PROPOSED 2024 EARTHSCOPE-OCEANS DATA POLICY

EarthScope-Ocean stations rely on financial support from individual organizations and funding agencies. As such, specific organizations may have different requirements, and as such, individual organizations may determine their own data release policy.

Some partners in EarthScope-Oceans have adopted a very **Progressive Data Policy** which states:
Data from at least 10% of an ESO partner's MERMAIDS should be released without any delay other than the time required for data curation. Data from all other stations will be released two years after recording.
EarthScope-Ocean Partners that have adopted the Progressive Data Policy include:
- Géoazur, (France)
- Princeton University, (US)

Specific Data Policies in place by other EarthScope-Oceans Partners
- Stanford (US)
  - (insert specific policy here)
- Jamstec (Japan)
  - (insert specific policy here)
- EOST (France)
  - (insert specific policy here)
- Sustech (China)
  - (insert specific policy here)
- University of Sao Paulo and Observatorio Nacional of Brazil (Brazil)
  - (insert specific policy here)

Data from EarthScope-Oceans shall reference ESO data using the International Federation of Digital Seismograph Networks (FDSN) DOI 10.7914/SN/MH

THANK YOU