

FDSN Framework Proposal for a GeoCSV format for rapidly changing parameters not handled well in SEED 20 August 2021

Proposal Phase – Community Consensus Phase Type B: new FDSN standard
Submitted by EarthScope-Oceans (ESO)
Tim Ahern, IRIS Emeritus and Joel Simon, Princeton

Background: As we all know, SEED and current StationXML, do not handle metadata that changes often in a compact manner. We propose a GeoCSV format to allow data producing centers to easily produce a CSV formatted file that captures most rapidly changing parameters. This model can be extended to accommodate additional rapidly changing parameters that may be specific to a given project.

This proposal is in support of a flexible and generalized approach that in theory can accommodate most if not all such parameters.

Current Status: Following upon a concept presented at the Montreal IASPEI meetings held in 2019, the EarthScope-Oceans team, led by Joel Simon, developed a GeoCSV formatted collection of information that captured the rapidly varying locations for the MERMAIDS project. They identified not only positional information, but also sensor and timing information that they wished to incorporate and submit to the IRIS Data Center as a single file.

Using this approach, it was simple to modify the existing EarthScope-Oceans (ESO) workflows and these GeoCSV files have now been submitted to the IRIS DMC and are being managed in the MetaData Aggregator (MDA) system at IRIS (see <http://ds.iris.edu/data/reports/MH/>). This is the first time that IRIS has captured changing positional information in a standard way. Prior to this, most groups with data from moving stations simply created their own ad hoc spreadsheets that varied from deployment to deployment. Some of this positional information now seems lost or at least were not findable.

ESO chose to submit one GeoCSV file for every MERMAID instrument package but there is no specific requirement and GeoCSV files could contain all information for all stations in the network or other aggregations. Using the ESO approach makes it easier to submit updates on a single station basis as needed.

The Elements Used in the Proposed GeoCSV File Following the ESO Model

An example of a MERMAID (FDSN network code MH) is provided in Appendix A.

- GeoCSV file header. The first several lines in the proposed file contain information such as the version of GeoCSV, the date and time the file was created, any file specific information such as the position interpolation method used (in this case automaid v3.4.2 with relevant DOI), the delimiter and lineterminator characters. While some of these are ESO specific most are general and should be included in most GeoCSV files.

- #dataset: GeoCSV 2.0
 - #created: 2021-07-28T23:25:20Z
 - #automaids: v3.4.2 (<https://github.com/earthscopeoceans/automaids> (doi: 10.5281/zenodo.5057096))
 - #delimiter: ','
 - #lineterminator: '\n'
- Following the header lines, there are three rows that contain the parameters for
 - the units
 - the type of parameter (i.e., string, float, time)
 - the name of the parameter
- Data Section
 - **Element: Method Identifier (Measurement or algorithm or ...)**
 - ESO Method Identifier is either a measurement or an algorithm. ESO has a more complicated problem where the timeseries recordings take place underwater, at an unknown position, and an algorithm must be used to determine the position of the instrument when it was recording timeseries underwater. The algorithm uses the actual measurements surrounding the algorithm lines in the GeoCSV file. Most GPS equipped systems would only have “measurement” as the method identifier. We believe ESO is one of the more complex methods of determining positional information.
 - **Element: Time and SNCL Information**
 - Start Time
 - Network
 - Station
 - Location
 - Channel
 - **Element: Position**
 - Latitude
 - Longitude
 - Elevation
 - Depth
 - **Element: Sensor (optional)**
 - Sensor Description
 - Sensor Dip and Azimuth
 - Sensor Scale/gain
 - Scale Frequency
 - Scale units (e.g., Pa, m/s)
 - **Element: Timing (optional)**
 - Sample Rate
 - Time Delay
 - Time Correction
 - **Element: User Defined (optional)**

The user may provide any other element they wish to if they decide it is needed for a fuller understanding of the related time series information. At times the information may be redundant with data in StationXML but included for simplification of various workflows.

Usage: While the information in the GeoCSV file can be used by applications in a standalone mode if desired, we think that ultimately this information should be incorporated into standard StationXML. We believe the above Element Names would become tags for a high-level element and the subordinate identifiers would be included under their own tags within StationXML. The mapping should be straightforward and naming conventions could assure tag uniqueness.

Steps to Consider

The Proposed Procedure – follow the FDSN Framework

Community Consensus Phase

1. WGV should form a Proposal Review Team
2. Proposal Review Team should be cognizant of the example set by EarthScope-Oceans
3. An attempt should be made to clearly identify tag names in a rational manner to ensure uniqueness
4. WGV should vote on the Review team recommendations
5. If approved, the WGV recommendation should be sent to the Steering Committee to consider a formal adoption by the FDSN

Evaluation and Adoption Phase

6. If the decision is made to accept the standard, then
 - a. WGII should begin the formal definition of an extended StationXML Schema to capture the information in the GeoCSV file
 - b. FDSN WGII could identify tools and systems to create, manage, and disseminate the information in the extended standard
 - c. Existing FDSN web services should be modified to enable access to information originating in the GeoCSV file

End of Framework Proposal

Appendix A. An example GeoCSV file for FDSN network MH, Station P0008

Type	Time and SNCL				Position				Sensor				Timing			
	ISO_8601 datetime	unitless string	unitless string	unitless string	degrees_nord float	degrees_east float	meters float	meters float	unitless string	factor float	hertz float	unitless string	hertz float	seconds float	seconds float	seconds float
#field_unit MethodId/en	Start/Time	Network	Station	Location	Latitude	Longitude	Elevation	Depth	SensorDescri	Scale	Scale	Scale	Scale	TimeDelay	TimeDelay	TimeCorrection
Measurement	2018-08-05T MH	P0008	P0008	nan	-12.008233	-172.0231	0	0	0	MERMAIDHy	nan	nan	nan	0.00003	nan	nan
Measurement	2018-08-05T MH	P0008	P0008	nan	-12.006967	-172.01872	0	0	0	MERMAIDHy	nan	nan	nan	-0.000062	nan	nan
Measurement	2018-08-06T MH	P0008	P0008	nan	-12.0477	-172.01357	0	0	0	MERMAIDHy	nan	nan	nan	0.437377	nan	nan
Measurement	2018-08-06T MH	P0008	P0008	nan	-12.047684	-172.01369	0	0	0	MERMAIDHy	nan	nan	nan	0	nan	nan
Measurement	2018-08-06T MH	P0008	P0008	nan	-12.047584	-172.01425	0	0	0	MERMAIDHy	nan	nan	nan	0	nan	nan
Algorithm:au	2018-08-08T MH	P0008	P0008	0 BDH	-12.074427	-171.99651	0	1531	0	MERMAIDHy	-149400	1 Pa	nan	20	nan	-0.29087
Measurement	2018-08-15T MH	P0008	P0008	nan	-12.205566	-171.90376	0	0	0	MERMAIDHy	nan	nan	nan	1.643707	nan	nan
Measurement	2018-08-15T MH	P0008	P0008	nan	-12.2059	-171.90405	0	0	0	MERMAIDHy	nan	nan	nan	0	nan	nan
Measurement	2018-08-15T MH	P0008	P0008	nan	-12.207367	-171.9055	0	0	0	MERMAIDHy	nan	nan	nan	-0.000184	nan	nan
Measurement	2018-08-15T MH	P0008	P0008	nan	-12.207784	-171.90599	0	0	0	MERMAIDHy	nan	nan	nan	0	nan	nan
Algorithm:au	2018-08-16T MH	P0008	P0008	0 BDH	-12.231257	-171.89015	0	1527	0	MERMAIDHy	-149400	1 Pa	nan	20	nan	-0.242693
Algorithm:au	2018-08-17T MH	P0008	P0008	0 BDH	-12.255579	-171.86842	0	1521	0	MERMAIDHy	-149400	1 Pa	nan	20	nan	-0.523292
Measurement	2018-08-17T MH	P0008	P0008	nan	-12.2612	-171.86581	0	0	0	MERMAIDHy	nan	nan	nan	0.569366	nan	nan

Reference: <http://ds.iris.edu/data/reports/MH/>