

Minutes of EarthScope Oceans Data Committee

18 November 2021 17:00 UTC

Attendees: Joel Simon, Christoph Waldman, Tim Ahern, Frederik Simons (joined 1 hour into meeting)

Introductions: Since this was the first meeting of the Data Committee Joel and Christoph briefly introduced themselves and their past and current activities.

Current Status of ESO data in IRIS DMC

Data have been available at the IRIS DMC for many months. The most current is from August 21, 2021. Joel has more data and will forward it to the DMC soon. Joel plans to implement the mseed2dmc application so that the data can be transferred to the DMC when it is ready with little effort. He is working with Rick Benson and Un Joe at the DMC to enable this.

December 2021 update: Joel has successfully implemented a `miniseed2dmc` workflow to automatically and periodically deposit new data with IRIS. The effort was especially helped by Un Joe and Inge Watson at IRIS.

GeoCSV file format to carry rapidly changing metadata

MERMAIDs are not stationary but rather drift in with the ocean currents. A new file format called GeoCSV was developed to capture the rapidly changing metadata including variations in latitude, longitude, and depth (among others), for these moving sensors. Joel began by showing an example of a GeoCSV file and what the various fields are. Files in GeoCSV format are available for the MH (MERMAID) data at the IRIS DMC within the IRIS Metadata aggregator system. (ds.iris.edu/mda/MH). Joel asked the committee to consider removing some of the current fields in the GeoCSV file since they are either not needed or the metadata already exists in the StationXML metadata files. Specifically, Joel sought agreement to remove the Scale, ScaleFrequency, and ScaleUnits parameters from the ESO GeoCSV files. The entire committee was in agreed they could be removed.

Joel also requested that ESO consider adding a "Measurement:depth" row, in the form of, e.g. "Measurement:depth:<pressure_gauge_name>:mbar measurement." By having a way of capturing all measured depth values as MERMAIDs rise and fall, reproducibility would be enabled. It would be a single row for each depth measurement taken by MERMAID, not just the depth reported at the time an event is detected. It was also agreed the native unit of "mbar" of absolute pressure, as measured by MERMAID, would be reported, as opposed to the "meters" for depth, of which we assume 1 dbar = 1 meter.

Status of ESO effort in the FDSN

In August 2021 the virtual meeting of the International Federation of Digital Seismograph Networks took place. Ahern and Simon attended some of these sessions.

EarthScope-Oceans was admitted as a regular member to the FDSN with Tim Ahern as the ESO representative within the FDSN. Frederik Simons is the alternate representative.

At the Working Group V session, Ahern submitted a framework proposal to have the GeoCSV file used for moving station data or for variables that are rapidly changing in time. A team is being assembled presently and Joel and Tim, while not team members, will act as a technical resource for their deliberations. An estimate is that sometime in the first half of 2022 a recommendation will come out of the team's deliberations.

If accepted by WG V, the task will then be passed on to WG II that is responsible for the StationXML format. Ideally, WG II will incorporate the GeoCSV information into StationXML and it will become a native element of StationXML.

Seismic data as a possible Essential Ocean Variable (EOV)

(Frederik joined shortly after discussion of this topic began)

Christoph gave us a quick review of the Essential Climate Variable effort that exists. Similar to the ECV, there is now a concept of Essential Ocean Variables (EOV) being led by the Global Ocean Observing System (GOOS). EOVs are identified by GOOS Expert Panels based on:

- **Relevance:** The variable is effective in addressing the overall GOOS Themes – Climate, Operational Ocean Services, and Ocean Health.
- **Feasibility:** Observing or deriving the variable on a global scale is technically feasible using proven, scientifically understood methods.
- **Cost effectiveness:** Generating and archiving data on the variable is affordable, mainly relying on coordinated observing systems using proven technology, taking advantage where possible of historical datasets.

When EOVs are identified, a series of recommendations are created and disseminated by the Expert Panels, including what measurements are to be made, various observing options, and data management practices.

Christoph believes that there are benefits for ESO to join this effort and make all the measurements made by MERMAIDs EOVs. After some discussion it was decided that this is a good idea, and this will become a topic on the upcoming Steering Committee Agenda.

Exchange of data with CTBTO

Christoph has had recent discussions with the German representatives of the CTBTO. He introduced the topic of the CTBTO's Virtual Data Exploitation Centers (VDEC) as an organization where data could be shared. Tim indicated he knew about the VDEC but for non-states parties such as the scientific research community there are some limitations. First anyone using data from the VDEC must get a No Cost Contract with the CTBO and agree not to share data they receive through the VDEC with anyone else. So ESO should be cautious about this. VDEC would be a method thorough which

ESO could submit data, but it would not be a viable method to get data broadly disseminated.

Everyone felt that building a relationship with the CTBTO would be a good thing and the committee encouraged Frederik to add this as a discussion point for the next Steering Committee meeting.

Building the ESO Data System

Significant progress has been made in building a method of sharing the ESO data. Largely through the efforts of Joel, Guust Nolet, and staff at the IRIS DMC ESO data from all Princeton and GeoAzur MERMAIDS is now available at the IRIS DMC. A newsletter article has been published by the IRIS DMC related to this with authors being the ESO Data Committee members.

It is now essential to establish principles and requirements of the ESO data system. Tim presented some guiding principles that all ESO Data Centers should meet. After the committee deliberated two additional elements were added. The following shows the recommendations the Data Committee is making to the ESO Steering Committee for adoption.

Recommended Requirements for ESO data centers

1. ESO Data Centers must
 - a. Support FDSN Standard data formats (SEED) and inbound transmission protocols (SeedLink) and mseed2dmc
 - b. Support all FDSN Webservices for outbound data (e.g., Station, DataSelect, Event, and Availability)
 - c. Participate in the FDSN federated data centers effort
 - d. Routinely calculate data quality metrics following FDSN requirements (IRIS MUSTANG or EIDAWS-WFCatalog Service) for all MERMAID data they make publicly available.
2. Until these requirements are met, MERMAID data will go to Joel for processing and then be forwarded to the IRIS DMC
3. Items 1.a, 1.b, and 1.c and 1.d are fully supported at the IRIS DMC. Item 1.d is addressed in the IRIS MUSTANG system.
4. An ESO component wishing to manage and distribute their own data could purchase the SeisComp3 system for use in their centers. SeisComp3 is available from GEMPA in Potsdam, Germany (<https://www.gempa.de>). Using the EIDA Quality Tools (<https://www.orfeus-eu.org/data/eida/quality/metrics/>) item 1.d above could be addressed
5. ESO Data Centers must provide rapidly varying metadata such as position in GeoCSV format
6. Partners that wish to leverage the work done by Joel should consider funding for a dedicated EarthScope Oceans Data Collection Center. This concept should be promoted by the Steering Committee. This ensures proper format compliance before transmission to an ESO data center.

7. A single citable codebase must be maintained. As OSEAN manufacturers new instruments that collect new and/or different data, or an OSEAN member adds new features to their processing codes, those updates must be communicated to ESO at large and all codes should be merged, citable and be shown to run for all users. For example, as of Dec 2021 the “automaid” codelines have unfortunately bifurcated into a purportedly “official” but now out-of-date “EarthScope-Oceans” version (automaid v3.5, <https://github.com/earthscopeoceans/automaid>, doi: [10.5281/zenodo.5514229](https://doi.org/10.5281/zenodo.5514229)), and a conflicting OSEAN version written to handle a new “bin” file format (<https://github.com/oseanfro/automaid/>). Branching and merging from a common code root must be encouraged for long term software stability and data reproducibility.

Recommended ESO Data Release policy

The data committee strongly recommends ESO data release meets a single ESO Policy. Furthermore, the committee recommends the following.

Recommended Data Policy and Citation

- i. 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.
- ii. Data from all other stations must be released after two years.
- iii. For more details concerning early access to restricted data please refer to <https://geoweb.princeton.edu/people/simons/earthscopeoceans/data/metadata.html>.
- iv. Reference ESO data using DOI 10.7914/SN/MH.