



ICES
CIEM

International Council for
the Exploration of the Sea

Conseil International pour
l'Exploration de la Mer

DIG September Operational Meeting
21 September, 2020

Oceanographic data processing considerations

Oceanographic data managed within the ICES Data Centre (see <https://www.ices.dk/data/data-portals/Pages/ocean.aspx>) is currently allowed to be submitted in any well described format, and then subsequently mapped to the ICES Oceanographic Format (IOF). This model was conceived at a time when computing power, and computing storage were at a premium, and there was no internet. In line with the focus on attracting oceanographers back into ICES, and the advisory plan promoting FAIR data principle, it is proposed to bring the ingestion format into line with well described and supported community practices. This would also allow the ICES Secretariat to focus on automation of quality control and standardisation to community vocabularies. Of the 30 dataflows that ICES have documented for the Core Trust Seal accreditation, the oceanographic data flow is the only one accepting a non-specified format.

Actions from this document

- Agree to the change towards accepting the two recognised community based formats;
- Agree to the proposed timeline
- Nominate reviewers for the 'ICES' version of ODV format

Significant limitations with the current data flow process

1. Invisible format: There is no description on the ICES website as each submitter can provide their own format. This is contrary to other formats used in data ingestion portfolio (<https://www.ices.dk/data/tools/Pages/Data-formats.aspx>). This is problematic for data submitters, and for ICES data management, in terms of being able to document and describe the dataflow, ensuring key information is consistently provided, and for relating this to known community standards.
2. Data processing and Quality Control house of cards: the incoming data are handled by a suite of disparate legacy programs and require a great deal of human intervention. The legacy programs are not supported by newer operating systems, are hard to maintain, and would be costly to refactor.
3. Vocabularies: Current international accepted vocabularies are not specified/mandatory, which limits data interoperability with our stakeholders and users. As an example, the 4 character IOF country/ship codes do not conform to the platform vocabularies (hosted at ICES, [//vocab.ices.dk/?ref=315](http://vocab.ices.dk/?ref=315) and governed internationally) but can be either a true platform code or alternatively a data separator e.g. county code, call sign etc.
4. Best practice data fields: Only by email are key fields requested if the data provider does not supply them in their initial submission. An example are core parameters where method or unit may not be supplied. This is a manual process, and prone to gaps and error.

5. Metadata: Metadata is supplied inconsistently - such as station name/number, assessment units, project/program, provider etc., nor a CSR reference which is currently considered the oceanographic metadata directory.

Proposal for community based standard format

It is therefore proposed to consolidate the dataflow to only allow data submissions in two formats that are well recognised and well used by the ICES data submitters. The primary ‘default’ submission format would be:

ODV (to be reviewed by DIG) ¹

The secondary format, which is already in use by some oceanographic data submitters to consolidate different data types:

ICES Environmental Reporting Format (RF 3.2)²

A sub-group of DIG should review the SDN version of the ODV format, to determine the need for modifications to the mandatory/optional fields to satisfy data quality/checking requirements at ICES³ (for instance, having platform code as mandatory).

Timeline

Step	When
Notice to data providers	September 2020
Review of ODV format specification	December 2020
Data ingestion to accept two community formats, and existing legacy formats	January – June 2021
Data ingestion to accept only two community formats	July 2021

¹ <https://www.seadatanet.org/Standards/Data-Transport-Formats>

² https://www.ices.dk/data/Documents/ENV/Environment_Formats.zip

³ <http://ices.dk/data/tools/Pages/quality-control.aspx>