Comments on the recommendations made in the POGO Working note ‘Research Vessels contributions to the Global Ocean Observing System’

Taco de Bruin (Royal NIOZ)

Roemmich, Gould and Desaubies in their POGO Working note ‘Research Vessels contributions to the Global Ocean Observing System’, correctly note that research vessels collect high quality and very valuable marine data, but that this data is not always made available to the scientific community. They recommend a series of actions to be taken to improve this situation. Summarized, these recommendations call for submission of data to open systems, (almost) immediately after collection of the data. Though we welcome measures to improve the availability of marine data, these recommendations need careful consideration.

The leading principles of good data management practice in (marine) science are:

1. Adherence to the IOC data policy of ‘free, open and timely access to marine data’. ’Free and open’ are easily defined, ‘timely’ may differ from case to case.
2. Single final version
   Data managers work hard to avoid the situation in which one original dataset results in several, different(!) final versions. Otherwise it’s impossible for scientists to decide whether they have the best dataset available.
3. Proper acknowledgement
   It is of paramount importance that scientists are credited for the work done on processing and finalizing data. Proper reference to the data originator should always be made when data is used in new applications. POGO could and should play an important role in this matter. Currently, scientists are credited for the number and quality of their publications. They don’t get credit for finalizing and submitting a good quality dataset. A Data Set Citation Index system, set up by POGO, may be a way to overcome this situation.
4. Protection of the interest of the data collectors

The recommendations show a tendency towards requirements for operational oceanography: immediate access to CTD, XBT and TSG data for use in models, without much care or time for a proper calibration and quality control. In the current form, these recommendations may conflict with several of the leading principles as outlined above.

CTD downcast data are either core data for a certain cruise/project or non-core data. In the case of core data, the sensors must be calibrated with independent measurements. Usually, sensors will be sent back to the factory for after-cruise re-calibration. The whole post-processing procedure may take several months. It will thus not be feasible to submit core CTD data in near-realtime.

Submitting the data in raw or preliminary form is not a good option for two reasons: The (operational) data centres mentioned will give open access to the data (leading principle 1). So the data will be available to everybody for any purpose. This will lead very easily to the existence of multiple, non-identical versions from the same original
dataset (even or especially when ‘decimated’ datasets are submitted), making it for scientists almost impossible to determine the best dataset. It is in conflict with leading principle 2. It is also in conflict with leading principle 4 (protect the interests of the scientist) to make core data available before the scientist has had a chance to work with the data.

Non-core CTD data require a level of accuracy which can be obtained by using factory calibrations. So non-core CTD data can be processed during the cruise and can be in final form shortly after collection. The data should be quality controlled before being submitted to a data archive, to prevent the existence of multiple (different) versions stemming from one original dataset. Onboard processing and quality control require additional resources.

As opposed to the remarks by Roemmich et al., TSG measurements do need ‘added effort’. Calibration samples should be collected during a cruise and the results from those calibrations should be applied to the data during the post-processing phase. This requires time and resources. We are not arguing that TSG measurements should not be made for the reasons given above. We are merely arguing that care should be taken to obtain good quality TSG data.

The note by Roemmich et al. “deals only with generic physical data: temperature and salinity”. Yet almost all RV also measure meteorological parameters and ocean depth on a routine basis. We believe that these measurements fall in the same category as CTD, XBT and TSG measurements and that therefore POGO could easily address those as well. Most, if not all, RV are part of the VOS fleet. Surprisingly enough, just a few RV are also part of the VOSClim programme. This should be remedied. Many RV are equipped with continuous recording meteorological sensors, measuring at intervals as short as 1 second. Yet the VOS and VOSClim programmes can only handle a dataflow of 1 report per RV every 3 hours. Valuable surface meteorological data is being lost because there is no archive for this data. Perhaps POGO, in close collaboration with JCOMM, can address this issue.

The same applies to the depth sounding measurements by echo sounders or multibeam, which should be submitted to the GEBCO programme.

We are fully aware that this answer echoes (albeit in different words) many of the concerns as expressed by Paul Tildesley and Fred Stein, with which we completely agree. We hope that it will contribute to a discussion to further improve the recommendations to be adopted by POGO.

Taco de Bruin

Data Manager - Royal Netherlands Institute for Sea Research (Royal NIOZ)
Chair - Netherlands National Oceanographic Data Committee (NL-NODC)
Chair - Joint SCAR/COMNAP Committee on Antarctic Data Management (JCADM)
Co-Chair - IPY Data Policy and Management Subcommittee (IPY-SC on Data)