# Australian National Mooring Network

# Quality Control Summit

## CMAR Hobart 10-12 December 2013

### 1. Introduction

The third annual ANMN QC summit took a workshop rather than presentation approach with participants from QC teams both across and between IMOS facilities. The three days of the summit were divided into sessions based on data stream of interest to particular working groups. These included sessions on ADCPs, bio-optics, data streams from biogeochemical sampling, profiling CTDs and working with the Matlab toolbox. This working group approach is in accordance with IMOS moving from start up and development to maintenance of an operational system.

As always we also allowed for plenty of time to both formally in the facilitated meetings, and informally during breaks, discuss the issues, meet and network.

### 2. General categories of impacts on data quality

The QC teams identified three general categories that impact on the quality of the data:

- **Environmental**: bio-fouling, knock downs, vandalism, interactions with vessels/wildlife
- **Instruments**: firmware issues, drift, hardware issues such as floods, calibrations
- **Operations**: mooring failure, deployment/recovery failures

The summit identified that feedback is required between operational staff, QC teams and user scientists across these three general categories. It was also emphasised that it is important that QC teams do not initially focus onto any one of these categories but consider each for the particular observation. The use of manuals and checklists were also seen as key to reducing impacts on data quality.

### 3. Working groups across facilities

It became evident that each working group spanned multiple IMOS facilities. For example both the NRS project of the ANMN and the PULSE project in ABOS collect biogeochemical samples that needs to be QC'ed and stored in an easily accessible database. The following overlaps between facilities were identified:

- **Moored CTD** – ANMN, ABOS, Sat Cal Val
- **Profiling CTD** – ANMN, ANFOG, ARGO
- **ADCPs** – ANMN, ABOS
- **BGC** – ANMN, ABOS, SOOP (CPR)
- **Bio-optics** – ANMN, ABOS, ANFOG, Ocean Colour

The use of the ANMN's Matlab toolbox to parse and handle data across various facilities was also of interest as was the development of deployment and the BGC databases. One option for future work would be to consolidate both data QC and data handling, leading to an IMOS QC summit rather than an ANMN QC summit.

## 4. Data labelling

A priority that emerged from the summit was developing a data standard so the level of QC undertaken was clear.

Four proposed levels of data labelling were discussed:

1. Raw or FV-01
2. FV00
3. FV01
4. FV02

The first of these, called 'raw' or FV -01, is data directly from sensors that has not been parsed into NetCDF. This data type would be held by the operators, with the exception of automatically processed and telemetered near real time data which will be delivered to eMii. The second, FV00 is data parsed by the Matlab toolbox into NetCDF with metadata from the deployment database but no flags, which is then provided to eMii. Third, was FV01, which included QC flags from the essential QC tests. The FV01 data would be the mainstay of the facilities datasets and would be set up in a manner that allows for batch repossessing by eMii. The final level, FV02 includes both essential and discretionary QC flags and also overview from a subject matter expert. This final FV02 level was considered a data product.

## 5. Essential vs. discretionary QC

There was considerable discussion regarding setting of boundaries between 'essential' QC and discretionary 'QC'. The general consensus was for QC to be lenient so that 'good' data was not labelled as 'bad'. This may be particularly important for interesting variability which could be labelled as 'bad' data as it is anomalous. In particular, automatic or technical QC systems need to be very lenient as they cannot place into context data in the same manner as subject matter experts (See Morello et al submitted 2013). From this work and experience, essential QC was seen as what can be provided by technical and operational staff. Discretionary QC was seen as requiring at least a support scientist who is supervised and in direct and ongoing contact with senior scientists who are subject matter experts. The outcome from the discretionary QC would be seen as a data product. Based on the current tests within the Matlab toolbox a cut off level between essential and discretionary was suggested as follows:

Essential QC Matlab toolbox (FV01)

1) Magnetic correction in true north
2) Pressure and height above sensor
3) Impossible date
4) Impossible location
5) In and out water test

6) Side lobe test
7) Global range test

QC – discretionary tools in the Matlab toolbox (FV02)

1) Constant depth from re-gridding
2) Time drift
3) Regional range test
4) Impossible depth
5) Spike test
6) Rate of change
7) Stationarity test
8) Climatology test
9) Teledyne test – goal to include into essential as threshold set QC

Others

10) Inter instrument comparisons
11) BGC to sensor comparisons
12) Remote sensing comparisons
13) Model comparisons

The difference between essential and discretionary is also of importance for reprocessing of data. As the discretionary QC produces a bespoke product this data is no longer easily applicable for automatic reprocessing of bulk datasets. Essential QC data can thus be reprocessed centrally by the data centre, while data that has undergone discretionary QC will need to be reprocessed by QC teams.

## 6. Operational QC

The best QC system is one that is fully integrated into a research team, with technicians, QC data support scientist and science users – both research and principle scientists - in close contact to discuss data issues.

It is also best practice if those involved in the first or essential QC system participate on cruises and are operational people. Close working relationship with deployment, mechanical and electronics technicians to discover and document obvious failures of instruments cuts down on operational, instrument and environmental QC issues that occur in the field. For instance:

- Mislabelling of data
- Placement of instruments up-side down or in the incorrect order
- Wire stretch
- Instrument slips
- Photographs of bio-fouling
- Capture in metadata redesigns ‘on the fly’

## 7. Conclusions

Checklists, procedures and review of data and methods can address many of these QC issues. Manuals, checklists and procedures should be loaded onto the ANMN homepage and advertised through Marine Matters and the working groups.

Both this report and presentations from the 2013 ANMN QC summit will be loaded onto the ANMN page on the IMOS website.

It is important to schedule the next meeting and keep momentum going. As data QC is very much across facilities an IMOS QC summit rather than an ANMN QC summit should be considered.

The time and costs of automatic reprocessing of data by eMii needs to be determined. This will be closely related to the extent of QC undertaken on the data.

QC above a certain level should be considered as a data product. The use of more advanced QC procedures if undertaken either automatically or by non subject matter experts may flag interesting variability as 'bad' data.

A common template for cruise and or data reports needs to be developed.