IMOS Integrated Marine Observing System

Quality control and data labelling

Marty Hidas

IMOS / eMII

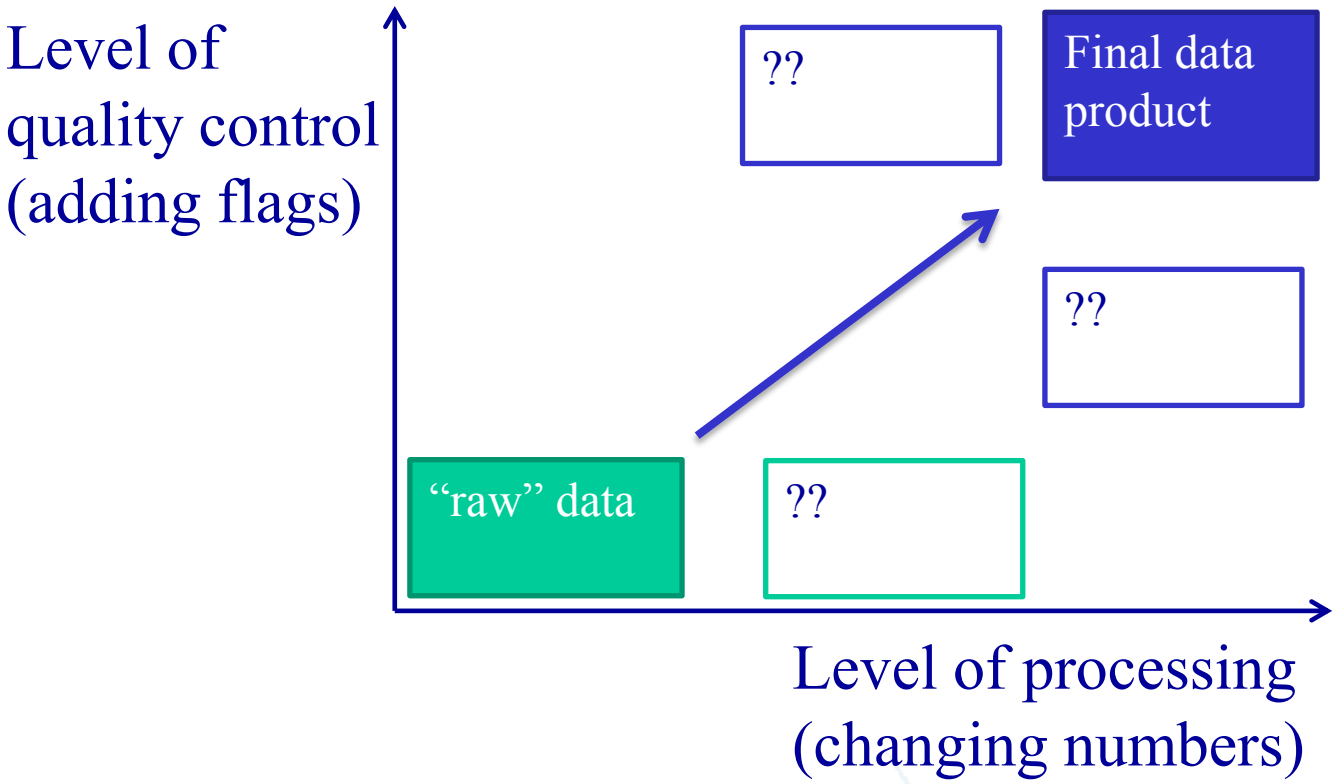
Australian Government
Department of Industry, Innovation
Science, Research and Tertiary Education

IMOS is supported by the Australian Government, through the National Collaborative Research Infrastructure Strategy and the Super Science Initiative

Currently we use two labels:

- FV00 – “Raw data”
  - "Raw data is defined as unprocessed data and data products that have not undergone quality control. The data may be in engineering units or physical units, time and locations details can be in relative units and values can be pre-calibration measurements. Level 0 data is not suitable for public access within IMOS."
- FV01 – “Quality Controlled Data”
  - "Quality controlled data have passed quality assurance procedures such as routine estimation of timing and sensor calibration or visual inspection and removal of obvious errors. The data are in physical units using standard SI metric units with calibration and other routine pre-processing applied, all time and location values are in absolute coordinates to agreed to standards and datum, metadata exists for the data or for the higher level dataset that the data belongs to. This is the standard IMOS data level and is what should be made available to eMII and to the IMOS community."

Two independent concepts:



- Use separate labels for them
- Also separate from
  - File format
  - Real-time vs. Delayed mode data

What do we mean by “quality controlled”?

- In/out of water test?
- + spike test?
- + gradient test?
- + climatology test?
- + XYZ automated test?
- + manual checking by an expert?

At which point do we start calling it “quality controlled”?

Beyond that, how can we indicate we’ve done more?

Clearly, need a way to indicate different levels of QC.

Suggestion

For QC:

0 = No QC (not the same as “raw”)

1 = Automated QC (IMOS standard set)

2 = Supervised QC (based on written procedures, additional to automated QC)

For processing level

0 = “raw” data, as read out of the instrument

1 = calibrated data (convert measured values to physical parameters)

2 = data products

Thoughts?