IQuOD

International Quality-Controlled Ocean Database

2nd Annual Workshop Report

June 4-6, 2014
NOAA, Silver Spring, USA.

http://www.iquod.org

Editor

Rebecca Cowley, CSIRO Marine and Atmospheric Research, Australia

Bibliographic Citation

IQuOD (International Quality-controlled Ocean Database) 2nd Annual Workshop Report, July, 2014.

Table of Contents

Workshop Summary 5

1. Setting the scene – current project structure, workplan and progress made in the last year. 6

1.1 Project aims and structure, Catia Domingues 6

1.2 Recap on goals for Auto QC, Manual QC, Data aggregation and Data assembly groups, Ann Thresher 7

1.3 The scientific implementation plan and plans for CLIVAR endorsement, Matt Palmer 9

1.4 Session 1 discussion. 10

2. Auto QC group benchmarking results 10

2.1 Auto QC group, Rebecca Cowley 10

2.2 Description of AOML tests, Francis Bringas. 11

2.3 Description of NODC tests, Tim Boyer 11

2.4 Description of CSIRO tests, Rebecca Cowley 12

2.5 Description of EN QC, Simon Good 12

2.6 ICDC Automated QC procedure, Viktor Gouretski 13

2.7 Auto QC testing results, Ann Thresher and Viktor Gouretski 13

2.8 Discussion on automated QC tests 13

3. Invited talks from data users and producers 14

3.1 Operational SI Climate Forecasting Data Requirements: Timeliness, Distribution, Types, QC, David W. Behringer NOAA/NCEP 14

3.2a Current status of Japanese XBT data reconstruction and related activity, Masayoshi Ishii 14

3.2b Progress in Auto QC by MIRC, Toru Suzuki, Marine Information Research Center, Japan 15

3.3 Quality Control for models/reanalysis, James Carton and Gennady Chepurin (University of Maryland) 15

3.4 Observing the Southern Ocean with the help of elephant seals. Fabien Roquet 16

3.5 QC methods at Coriolis/IFREMER, Christine Coatanoan 16

3.6 How might IQuOD interface with climate Model Intercomparison Projects (MIPs) and obs4MIPs? Paul Durack for Peter Gleckler. 17

4. User requirements and interactions 17

4.1 User requirements gathering - experience from the ESA Climate Change Initiative project on SST, Simon Good 17

4.2 Discussion on user requirements 18

5. Data types and flows 18

5.1 Data types, Viktor Gouretski 18

5.2 Discussion on data types 18

5.3 Data flows, Tim Boyer 19

5.4 Discussion on data flows 20

5.5 Ocean Data Standards: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. Hernan Garcia. 20

5.6 File content and data flagging, Ann Thresher 21

5.7 Discussion on file content and data flagging 22

6. Specific areas of work 1 23
6.1 Assigning uncertainties to individual observations: experience froom building HadIOD, Simon Good 23
6.2 Discussion on uncertainties 23
6.3 Data aggregation, Tim Boyer 24
6.4 Discussion on data aggregation 25
7. Specific areas of work 2 25
7.1 Intelligent metadata, metadata recovery and bias adjustment, Simon Good 25
7.2 Discussion on metadata, bias adjustments. 25
7.3a Manual QC, Ann Thresher 26
7.3b Crowd Sourcing for IQuOD manual quality control? Matt Palmer and Philip Brohan, Met Office Hadley Centre 26
7.4a Discussion on Manual QC 27
7.4b. Discussion on Crowd Sourcing 27
8. Refining the project structure and plan 1 28
8.1 and 8.2 Action items and discussion on action items 28
8.3 Next meeting: 28
9. Refining the project structure and plan 2 28
9.1 Funding 28
9.2 Working group goals and membership 30
9.3 IQuOD group structure review 30
9.4 Further discussion 30
10. Communication plans and close 30
10.1 and 10.2 Communication and webpage discussion 30
10.3 Other business: 30
10.4 Review action items 31
10.5 Close 31
Appendix 1. Workshop Agenda 32
Wednesday June 4 32
Thursday June 5 35
Friday June 6 36
Appendix 2. Participant list. 38
Appendix 3. Working group membership and tasks 2014-2015. 40
Data formats, uncertainty, flags 40
Aggregation working group: 40
Auto QC working group: 40
Manual QC working group: 41
GDAC working group: 42
Appendix 4. Action Items. 43

Workshop Summary

The 2nd IQuOD (International Quality-Controlled Ocean Database) meeting was held at NOAA, Silver Spring, US from 4-6 June, 2014. The project's structure and goals were refined and clarified and it was clear that a published meeting report, a scientific report on the project and the scientific implementation plan are essential for seeking funding and in-kind support. One of the key items for the meeting was to review the automatic QC benchmarking test efforts from the previous year. Some progress was made on these tests, during the next year, the benchmarking tests will be refined, finalised and reported on. The manual QC group was given several tasks for the coming year to begin comparison and clarification of manual QC methods, including a realistic estimation of the cost of manual QC and the goal of a manual QC workshop prior to the next IQuOD general meeting. A new task group to investigate file formats, uncertainty estimation methods and flagging methods was formed and will move forward with investigating existing schemes that will suit IQuOD. Funding from all countries is limited for the present, but in-kind support, allocation of hours within existing projects and working with existing projects are workable methods for now. Support from IOC subgroups is being sought and funding schemes being further investigated.

Major outcomes from the second IQuOD workshop include:

1. Review of the IQuOD structure, task group memberships and tasks.
2. Inclusion of a formats, uncertainties and data flagging task group led by Simon Good (Met Office, UK).
3. Efforts to clarify the mission statement for IQuOD and focus on the user requirements.
4. Clarification of the data types and instrument types that will be tackled initially in the project.
5. Planned publication of the scientific implementation plan, a meeting summary, scientific report and report on the Auto QC comparisons.
6. Action items as listed in Appendix 4.

1. Setting the scene – current project structure, workplan and progress made in the last year.

1.1 Project aims and structure, Catia Domingues

The most challenging application of subsurface ocean temperature/salinity observations (two of the Essential Climate Variables) is understanding climate variability and change (e.g., Earth's planetary balance, hydrological cycle and sea level), as it demands the highest data quality, completeness and consistency. Particularly, long-term high quality historical records are required to separate anthropogenic influence from natural climate modes of variability (e.g., ENSO, NAO, IOD, SAM, PDO, AMO, etc). Subsurface ocean temperature and salinity are also widely used to either evaluate/constrain, initialize, or are assimilated into numerical models to investigate physical mechanisms and causes of past and current changes, and to predict/project future changes in our climate. There is, however, an urgent need to maximize the full potential of tens of millions of historical temperature (and salinity) profiles – collected since 1900s (or before) – to a vast range of climate-related research, applications and services of societal benefit.

To overcome the above difficulties, a new, internationally-coordinated effort – IQuOD – is being organized by the oceanography community, along with experts in data quality and management, and in consultation with end users (e.g., climate modellers, metrics panel, ocean/coupled reanalyses (e.g., GODAE)) and the broader climate-related community.

IQuOD's goal:

The overarching goal of the IQuOD initiative is to produce and to freely distribute the highest quality, complete and consistent historical subsurface ocean temperature global dataset (to its maximum extent), along with (intelligent) metadata and assigned uncertainties and a number of downstream added-value products. In the future, plans include extension of a similar effort to other subsurface ocean variables, such as salinity, oxygen and nutrients.

IQuOD's goal will be achieved by developing and implementing an internationally-agreed framework. No individual group has the expertise/resources to complete the above task. International coordination/cooperation is essential to the success of the IQuOD initiative. By pooling expertise and resources into a single best practice community effort, the IQuOD initiative plans to achieve the best outcome over the shortest timeframe, at the same time to avoid duplication of human and infrastructure resources. Although internationally-coordinated efforts exist for the ocean surface and atmosphere-ocean observations, no similar effort has been undertaken for the historical subsurface ocean observations to this date.

Expected outcomes include:

- Development/implementation of international standard practices for automated/manual quality control of historical (and modern) temperature data.

- Free access to historical data (raw and interpolated products), (intelligent) metadata and uncertainties.
- Numerous downstream applications of the IQUOD dataset for Earth system/climate-related research and services of great societal benefit, including future CLIVAR (e.g., *Ocean Climate Indicators; Research Foci Initiatives*) and WCRP (*Grand Challenge on Sea Level Rise and Regional Impacts*) priority-related activities.
- Template for future efforts: great community interest in improving the quality, consistency and completeness of the historical salinity observations and other ocean variables.

Meetings summary: June 2013 - June 2014

CLIVAR-GSOP Coordinated Quality-Control of Global Subsurface Ocean Climate Observations: International Quality-Controlled Ocean Database (IQuOD) inaugural meeting
12-14 June 2013, Hobart, Australia

IQuOD town hall session in the 2014 Ocean Sciences meeting,
23-28 February 2014, Honolulu, USA

2nd IQUOD workshop
4-6 June 2014, Washington DC, USA

IQuOD discussion session in the (Second) Meeting of the Joint IODE-JCOMM Steering Group for the Global Temperature-Salinity Profile Programme
17 - 20 June 2014, Oostende, Belgium

General recommendations and developments after 1st workshop/Town hall meeting (and prior to 2nd workshop):

- Establishment of 4 task teams: Automated QC (Bec Cowley); Manual QC (Ann Thresher); GDAC (Tim Boyer); Aggregation (Catia Domingues)
- Establishment of an extra working group: flagging/uncertainties
- Engagement with groups (ad hoc)/specify level of involvement
- Tighter engagement with end users community
- Project strategy: scientific/implementation plan and endorsements required
- Identification of funding opportunities & development of applications
- Planning corporate structure/image, website and communication

1.2 Recap on goals for Auto QC, Manual QC, Data aggregation and Data assembly groups, Ann Thresher

IQUOD established 4 working groups at the first meeting a year ago:

Aggregation working group – to pull datasets together and ensure required metadata is gathered, Catia Domingues: group leader

Members: Tim Boyer, Gustavo Goni, Viktor Gouretski, Nathan Bindoff, Steve Diggs, Jim Swift,Marty Hidas, Sebastien Mancini, Guy Williams, Uday Bhaskar, Toru Suzuki, Sergey Gladyshev

Goals:

- Provide an inventory of the problems for available historical data/metadata
- Provide a list what data/metadata we don't have – can we get this?
- Begin acquisition of readily available sources of QCed/flagged data into WOD (Hydrobase/ENACT, etc).
- Begin engaging parties in acquisition
- Begin provision of intelligent metadata

Auto QC working group – assess various techniques and pick the best, then apply these to the aggregated dataset, Rebecca Cowley: group leader

Members: Ann Thresher, Tim Boyer, Jeff Dunn, Shoichi Kizu, Gustavo Goni, Viktor Gouretski, Guy Williams, Matt Palmer, Simon Good.

Goals (short term):

- Work towards a consensus on auto screening tests
- Do this by testing different auto screening methods on two (or more) datasets that have been fully QCd
- Get one person (student?) to assess the performance of each auto qc screening
- Re-grouping to discuss the results

Goals (longer term):

- Gather together a group of experts in individual instrument types to develop a scheme for assigning error estimates (Bec to begin coordinating, but might be better to engage someone else for this job)
- Provide a clear statement of what the AutoQC group would like to achieve
- Document a standard list of tests

Manual QC working group – devise QC software and train operators to apply manual QC to the Auto QC'd dataset, Ann Gronell Thresher: group leader

Members: Shoichi Kizu, Matt Palmer, Gustavo Goni, Simon Good, Uday Bhaskar, Ping Robinson, Alison Macdonald, Rebecca Cowley, Molly Barrenger, Lisa Lehman, Giles Reverdin, Alexandro Orsi, Guy Williams

Goals (short term):

- Define codes, flags and tests to be applied. Some of these will be instrument specific, others will be general.
- Establish rules for quality codes (e.g., you cannot apply a wire break to a bottle, XBT data cannot have good data below bad except in very specific

circumstances), decide whether we can change data (interpolate spikes? Move data subject to premature launch failure?)
- Document codes and procedures and provide detailed instructions for QC.
- Establish a group of experts to advise on regional issues or issues specific to a particular instrument type
- Provide an estimate of costs of doing manual QC for funding applications.
- Investigate Crowd Sourcing

Goals (long term):

- Set up a group to monitor and assess the global manual QC as it is returned to the master database.

GDAC working group – define and establish a GDAC as a home for the end result, Tim Boyer – group leader

Members: Nathan Bindoff, Steve Diggs, Jim Swift, Simon Good, Susan Wijffels, Matt Palmer

Goals:

- Coordinate all task teams
- Establish a steering group (Director, Co-chairs: provide clear directions)
- Steering group role: ensuring delivery of what we promised (quality/consistency) and managing the integrity of the workflow
- Investigate possibilities for GDAC locations and define GDAC roles and resources needed (NODC, Coriolis, ?)
- Define & track metrics
- Establish documentation/products to meet end user requirements (FAQs, flags/errors/data products/ documentation/traceability/manuals)

1.3 The scientific implementation plan and plans for CLIVAR endorsement, Matt Palmer

The value of seeking CLIVAR, and potentially other endorsements, was discussed. The scientific implementation plan is an essential document and will help to focus group plans, organisation and help promote successful funding proposals. An overview of the plan was presented and discussed, with the following points emerging from discussions:

- Need to clarify the overarching goal of the project - i.e. working towards a "best" profile database. There is merit in having additional variables (e.g. salinity) in scope, but say that we are starting with temperature. These additional variables will need to be touched upon in the science background/legacy.
- It was suggested that a meeting summary/discussion be written up for EOS. In addition, we should consider publishing a BAMS article based on the science background and potential legacy of IQuOD. **Action item.**

- It was agreed that a first draft of the implementation plan will be prepared by Matt Palmer and Catia Domingues and circulated no later than 30th June 2014. At this point the document would be open for comments/review by other members of the IQuOD project. **Action item.**

1.4 Session 1 discussion.

The group was asked to reflect on the previous year and discuss any ideas on what was done wrong and where we can improve.

IQuOD goals should state specifically and clearly that the aim is a QC'd historical dataset. We also want to establish the processes to get to the product. There was some discussion on whether we want to develop value added products such as climatology/gridded products - are there opportunities for funding here? The general consensus was that climatologies and gridded products could be easily made by others as part of funded projects, but our focus is on the observational level data.

We need to make clear to users the differences between WOD and IQuOD in the scientific implementation plan, how they will exist together and what purpose each serves.

Any software products should be distributed to help people do the manual QC.

Are we flagging observations or just assigning uncertainties? A combination of both will be needed. We should try to use existing systems (formats, quality control flags, uncertainty estimate methods) as much as possible.

How can we make IQuOD attractive to funding managers and justify costs?

- IQuOD should target a broad audience e.g. GODAE, CLIVAR etc.
- IQuOD will make economic efficiencies and scientific improvements.
- We need to clearly articulate why we need to make investments e.g. GDACs.
- We need to make clear that existing projects such as WOD and GTSPP have different objectives and do not provide the same level of QC that IQuOD will.

Action: We need letters of support from agencies such as CLIVAR, US CLIVAR etc.

2. Auto QC group benchmarking results

2.1 Auto QC group, Rebecca Cowley

At the first IQuOD meeting, the Auto QC working group was tasked with comparison of automatic quality control methods used by institutions participating in IQuOD. Three scientifically QC'd datasets were made available for the testing (the CSIRO Quota dataset from the Indian Ocean, ~114000 profiles on observed levels, a North Sea dataset from ICDC, ~500000 profiles on standard levels and a seal tag dataset provided by Fabien Roquet from the Southern Ocean, ~10000 profiles).

Using these three datasets with QC flags removed, each participating group was asked to apply their automated QC methods and return the results to Ann Thresher for comparison. The project is partly underway, but we have had issues with data formats and getting a complete result before the workshop. The project will continue this year.

2.2 Description of AOML tests, Francis Bringas.

The automatic quality control (AQC) procedures in use at AOML for XBT profiles in real-time was tested against three data sets (QUOTA, North Sea and Seal Tag) in order to assess the number of good and bad profiles detected by this system. AOML's AQC include the following tests:

- Gross check
- Constant value
- Spike
- Vertical Gradient
- Climatology
- Analysis
- Date
- Location
- Depth

While AOML's AQC procedure typically help identify as good and submit to the GTS ~95% of all profiles received at AOML in real-time, the results of the IQuOD test showed that a large number of good profiles were flagged as bad according to one or more of the tests listed above: 73% in the QUOTA data set, 68% in the North Sea data set, and 58% in the Seal Tag data set. The main reason for this is believed to be the absence of Climatology or NCEP Weekly Analysis data in certain regions, as well as several restrictions of AOML's AQC procedures. For example profiles older than 1997 are automatically flagged as bad as well as profiles with less than 5 points. However more information about the data sets tested is needed in order to better assess these assumptions.

It was recommended to repeat the QC test with a reduced and uniform data set and with modified procedures in order to better identify strength and weakness of each test in the search for the best procedures according to IQuOD requirements.

2.3 Description of NODC tests, Tim Boyer

This talk was limited to the World Ocean Database (WOD) quality control (qc) tests, as that is the qc test set which is being analyzed as part of the IQuOD auto-qc comparison.

All qc (automatic and manual) for the WOD are performed with a specific purpose in mind - the calculation of the World Ocean Atlas climatological mean fields and ocean heat/salt content calculations. The tests deliberately flag data which may be good observations, but do not represent a long-term mean or (for heat/salt calculations) a short-term large scale pattern.

The automatic qc consists of min/max range checks, spike checks (excessive gradients and inversions), stability checks (temperature+salinity), and standard deviation outlier checks. In total approximately 6% of all data are flagged in one or more of the tests.

QC flags values carried with the observations represent particular qc tests failed, not a verdict on the overall quality of the data.

Manual tests employed in calculation of climatologies and heat/salt content are included as qc information in the carried flags and enhance the usability of the data and reproducibility of results. Manual tests employed when uploading data to the WOD (speed made good checks, duplicate checks, etc.) also enhance the value of the data.

NODC is actively working to improve quality control procedures, particularly standard deviation outlier checks in regions with non-Gaussian distributions of observed values.

2.4 Description of CSIRO tests, Rebecca Cowley

The testing of CSIRO automated QC routines has been delayed because of software re-writes. The testing is not complete, but some results were available for the comparisons. The automated QC from CSIRO includes tests for:

Independent profile tests:

- Spikes
- Gradients
- Constant temperature
- Wire breaks (XBTs)
- Surface spikes (XBTs)
- Missing value clean-up for bottles

Climatology tests (using CARS):

- Temperatures outside 3x stddev of climatology.
- Integrated temperature divided by depth
- Mixed layer tests (compared to surface, and rate of change)
- Gradients.

It was recognised that the CSIRO tests do not include whole profile tests such as checks on location, date/time, lat/long.

2.5 Description of EN QC, Simon Good

The automatic quality control (QC) checks used to produce the Met Office Hadley Centre's EN4 dataset were applied to the IQuOD test datasets. The QC tests are: rejection of bathythermograph data shallower than 4 m and deeper than 950 m; profiles with latitude and longitude of exactly zero degrees are rejected (although this was not done for the IQuOD data); temperature values are checked to make sure

they are not ridiculous; a bathymetry dataset is used to check that the profile location is not on land; profiles are checked for spikes, steps and constant values; there is a track check; check of stability in the profile, that the measurement depths are in the correct order and not ridiculous; a Bayesian check against a forecast of the ocean state produced by persisting anomalies of an objective analysis from the previous month; a buddy check and a multi level check that rejects a whole profile if half the levels or more were rejected. For the EN4 dataset the multi level check is the main cause of a rejected profile while the background and buddy checks are the main cause of levels being rejected.

2.6 ICDC Automated QC procedure, Viktor Gouretski

The automated QC-procedure developed in ICDC (Integrated Climate data Center, University of Hamburg) was presented with the detailed description of the distinct data checks and the statistics of the quality procedure applied to each of the test datasets. The results of the QC procedure application to all main types of the WOD13 were also presented, with the graphic software allowing an easy overall quality assessment of the data.

2.7 Auto QC testing results, Ann Thresher and Viktor Gouretski

Three highly QCd test datasets were constructed from master datasets, representing a subset of the master sets with data reserved for future validation. These data sets were then processed by 5 centres who have auto QC code available. Others might be added as well.

Results were sent to CSIRO for comparison with the master quality flags. The analysis was complicated by the fact that every test data set and every test system used a different format.

The results depended both on the QC system being run and the input test dataset. The effectiveness of the tests varied from catching as little as 6% of the bad data known to be in the test dataset to catching as much as 65% of the bad data. One consequence of identifying the bad data is that you necessarily misidentify good data as being suspect. The systems that were best at identifying bad data also caught the highest percentage of good data.

The total number of profiles to be manually QCd depends on the combination of these two results. At this point, the optimal combination would require manual qc of between 20 and 50% of the data but would identify less than 60% of the known data of questionable quality. We now need to work to identify the individual tests that worked best and yet minimize the number of profiles we need to consider in Manual QC, as well as improving the tests to identify a larger proportion of the bad data that is present.

2.8 Discussion on automated QC tests

It is clear that we need to do more work. The test datasets should be refined, and we need to get more detail on the QC tests that have been applied to the data by each group. The tests that flag data need to be identified in the datasets returned to Ann.

Paul Durack suggested we look at using a model dataset (artificially created dataset) and run the autoQC over the top. Jim Carton, Peter Thorne might be able to assist with how to do this or if it is valuable. Matt suggested we could estimate the noise in models to get variability for the manual QC. We would need to use 0.5 degree resolution models. More detail would be required for this work.

Action: Write a paper on the AutoQC comparisons.

The AutoQC group tasks/focus:

- Quantify the manual QC tests done on each test dataset. Can we turn some of these into auto QC tests?
- Come up with next steps for auto QC group members.
- Look at incorporating density/T-S tests for QC of temperature.
- Provide failure rates for each test.

3. Invited talks from data users and producers

3.1 Operational SI Climate Forecasting Data Requirements: Timeliness, Distribution, Types, QC, David W. Behringer NOAA/NCEP

At NCEP we have two operational climate modeling systems, the Climate Forecast System 2 (CFS2) and the Global Ocean Data Assimilation System (GODAS). Each system produces ocean analyses using a 3-dimensional variational data assimilation scheme. Both systems rely on near real-time observations from the same sources: XBTs, CTDs, TAO, TRITON, PIRATA, RAMA, Argo, gliders, SST, SSS, SSH (TOPEX/JASON). The global observing system has changed dramatically over 25 years. When NCEP began model based climate forecasts we relied heavily on the TAO moored buoy array and XBTs and the observations, which sampled only the upper 500-750 meters, were largely confined to the equatorial zone and the northern hemisphere. There were very few salinity observations. Now we still rely on Tao and other tropical moorings although they are increasingly subject to budget constraints. We also now have the near global Argo array that provides both temperature and salinity down to 2000 meters and enables global ocean analyses with greater accuracy to greater depths. The Argo salinity data have proved particularly valuable in the equatorial zone leading to improvements in the zonal currents important in ENSO forecasting. In the immediate future we need to ensure the continuity of an affordable tropical observing system whether it remains based on moorings or evolves to a combination of moorings, Argo floats that spend less time at the surface and gliders. On a wider scale we face the challenge of maintaining and extending the global observing system in a way that will allow us to address new problems such as forecasting seasonal sea ice extent on a changing planet.

3.2a Current status of Japanese XBT data reconstruction and related activity, Masayoshi Ishii

The current status of the Japanese XBT data reconstruction was reported. We are ready to re-archive the data. The Japanese XBT data are rather in a good condition.

That is, most of meta data and data at every 1-meter depth are available at individual centers and institutes. According to the Tsurumi Seiki Co, Ltd (TSK), they sold 0.36 million probes so far. Meanwhile, the JODC stores 0.27 million TSK-XBT reports. Therefore about 0.1 million reports are thought to be unreported, This fraction occupies 25% of the total TSK XBT probes. In addition, TSK kindly provided us leaflets on MBT and the other BTs they produced. Each leaflet is 2-3 page document and pertly English translated. These will be shared with the IQuOD participants later. Moreover, our new climate study, a centennial ocean-atmosphere analysis with a coupled dynamical model was introduced. Assimilating historical ocean data is desired to reduce biases in the ocean model, that appear in the preliminary results.

3.2b Progress in Auto QC by MIRC, Toru Suzuki, Marine Information Research Center, Japan

Marine Information Research Center (MIRC) tried to improve the existing quality control procedures to QUOTA dataset. At first step the missing values such as -999, -99.9 or -273.15 were identified and assigned to missing value flag following recommendation of quality control flag scheme (IOC/2013/MG/54-3). Similarly whole same values in a profile, excessive outliers by broken wire of XBT were also identified. Outlier value or depth using measurement range by instrument were used and assigned to bad flag. Noise or spike values were assigned to questionable flag by excessive gradient and inversion check. About 6 % of samples were caught as error or questionable by above several steps of QC procedures effectively without existing range check in each basin as a function of depth, and header checks (date/time/position/ship speed/land-sea/maximum depth) will be also required for seeking more error or questionable profiles.

3.3 Quality Control for models/reanalysis, James Carton and Gennady Chepurin (University of Maryland)

In the first half of the talk we present a brief overview of the data sources and quality control (QC) procedures used in the Simple Ocean Data Assimilation analysis. The primary source of profile data is the World Ocean Database. Much of our QC procedure is legacy. We applaud the effort by IQuOD to provide a community-standard dataset. In the second half of the talk we discuss ways in which data assimilation can contribute to QC, for example by flagging observations whose values are inconsistent with the data assimilation background estimates (O-B differences). We illustrate this procedure showing results from a recent study (Giese et al., J. Clim., 24, 84-93, 2011) in which different BT bias correction algorithms are evaluated based on how they affect O-B differences. In our conclusions we argue for 1) allowing the reanalysis community to add observation flagging information to the database, 2) conducting the evaluation of temperature and salinity observations together (when available) to exploit information about T/S characteristics. Finally, if it is of interest to the IQuOD community we offer to construct a model profile sample dataset by sampling a high resolution ocean GCM simulation at the same spatial and temporal locations, depths and variables as the historical observations were

collected. Such a dataset can be used by the IQuOD community to conduct a variety of observing system simulation experiments.

3.4 Observing the Southern Ocean with the help of elephant seals. Fabien Roquet

Department of meteorology of the Stockholm University, Sweden
(fabien.roquet@gmail.com)

Over the last decade, several hundred seals have been equipped with conductivity-temperature-depth sensors in the Southern Ocean for both biological and physical oceanographic studies. A calibrated collection of seal-derived hydrographic data is now available, consisting of about 250,000 profiles. This includes extensive data from the Antarctic continental slope and shelf regions during the winter months, which is outside the conventional areas of Argo autonomous floats and ship-based studies. The seal data is quality controlled and calibrated using delayed-mode techniques involving comparisons with other existing profiles as well as cross comparisons similar to established protocols within the Argo community, with a resulting accuracy of ±0.03^{o}C in temperature and ±0.05 in salinity or better. The value of these hydrographic data within the existing Southern Ocean observing system has been demonstrated by conducting two state estimation experiments using the ECCO state estimate machinery. Including seal-derived data substantially modifies the state estimate within and south of the Antarctic Circumpolar Current, improving overall agreement with independent satellite observations of sea ice concentration. The data offer invaluable new insights into the water masses, oceanographic processes and provides a vital tool for oceanographers seeking to advance our understanding of this key component of the global ocean climate.

3.5 QC methods at Coriolis/IFREMER, Christine Coatanoan

Coriolis, as part of the French operational oceanographic system, has been especially involved in gathering all global ocean in-situ observation data in real time, and developing continuous, automatic, and permanent observation networks. Based on the Argo project, series of automated tests are defined and applied in real-time. For the Global Ocean, on a daily basis, gridded objective analysis fields of temperature and salinity are produced using profiles from the in-situ real time database (Argo floats, GTS data, XBT, CTD and XCTD, Sea Mammal and Mooring) of the global in-situ center and used for checking the data consistency. In the case of the in situ data assessment, the residual from objective analysis are used to detect outliers that have gone through the automatic tests and anomalous data that are then checked by an operator with a visual quality control tool and eventually flagged as bad if necessary.

Coriolis contributes also to the quality of ocean data analyses by producing every year an updated qualified temperature and salinity data set, and performing a scientific assessment of this data set extracted from the Coriolis data base. The Coriolis dataset for Re-Analysis (CORA) contains in-situ temperature and salinity profiles from different data types (Argo, GTS data, VOS ships, NODC historical

data...). The latest release CORA4 covers the period 1990 to 2012. Several tests have been developed to ensure a homogeneous quality control of the dataset and to meet the requirements of the physical ocean reanalysis activities (assimilation and validation). Improved tests include some simple tests based on comparison with climatology and a model background check based on a global ocean reanalysis. Feedbacks are also provided by modelers. Visual quality control is performed on all suspicious temperature and salinity profiles identified by the tests, and quality flags are modified in the dataset if necessary. CORA4 is available on request through the MyOcean Service Desk (http://www.myocean.eu/).

3.6 How might IQuOD interface with climate Model Intercomparison Projects (MIPs) and obs4MIPs? Paul Durack for Peter Gleckler.

obs4MIPs is a project that has been embraced by the WCRP's Data Advisory Council (WDAC) as a mechanism to better connect a diverse community of observational experts (such as IQuOD) with the climate modeling community. One way this is accomplished is to leverage the ~15 years of work to describe and organize climate model data, and now, satellite data and reanalysis. More information can be found at:

https://www.earthsystemcog.org/projects/obs4mips. Thus far, obs4MIPs is all about satellite data, but efforts are underway to expand this.

4. User requirements and interactions

4.1 User requirements gathering - experience from the ESA Climate Change Initiative project on SST, Simon Good

The European Space Agency (ESA) Sea Surface Temperature Climate Change Initiative (SST CCI) project was set up to try to realise the full potential of the satellite SST record for climate research. A first step was to produce a User Requirements Document (URD). User requirements gathering can take different forms. A particular group of users may be targetted and products designed to closely match their needs. Other users will likely find the products useful even though they were not specifically designed for them. Alternatively, an attempt can be made to capture the requirements of all users, which gives a comprehensive resource but it may be difficult to satisfy everyone. For the SST CCI project the latter approach was taken. A variety of methods were used to gather requirements: review of documents produced by other projects/institutions, discussion sessions with users, asking for lessons learned from other projects, an online questionnaire and asking 'trailblazing' users to trial the data and report back. All methods were useful but the questionnaire yielded the most information. It is recommended that a questionnaire be kept short and focussed. Questions should be chosen carefully and might be aimed at steering product design or may be aimed at providing a target that the products can be assessed against to see if they meet requirements. However, in some cases the limits of the data and/or the reference data against which the products are assessed may mean that a requirement is not possible to achieve or not possible to

demonstrate that it is achieved. Once data products are produced it is valuable to engage users by getting them to try the data and feed back their experiences.

4.2 Discussion on user requirements

How do we get more information on user requirements and can we make them part of the scientific implementation plan? Should we make a questionaire? At this stage, we should just target users to get their requirement feeback. We can invite specific users to give us direct feedback on what they want – maybe a short seminar series? Catia suggests get Karina von Schuckmann involved from a user perspective.

A feedback form on the website could help too. Look at the ESCOG website – it is a project home with interaction between users and producers. We need to focus on communities: Data assimilation, metrics (MIPS), models and data combined users.

Paul suggests that we open a conversation between IQUOD and MIP users. Matt suggests that we might get distracted by this. We should focus on profile QC at this stage, not a gridded product.

Tim Boyer, Jim Carton and David Behringer said that they wouldn't use the QC from IQUOD in their products. They would continue to use their own QC because they have a history, and specific requirements for their products. It may be that we are going to have to provide some gridded product. Products or papers? Papers are connected to products.

Action: Find out abou the pitfalls experienced from other programs: for example ICOADS. Talk to people next week at CLIMAR & GTSPP. Bec, Catia.

Action: A timeline needs to be incorporated into the implementation plan.

How do we incorporate a timeline without knowing what we are funded for? Janet suggests that you make a timeline assuming that you have funding and go from there.

5. Data types and flows

5.1 Data types, Viktor Gouretski

An overview of the main data types (bottle, CTD, Argo, APB, moored buoys, XBTs, MBTs) was given and the main problems specific to each data type outlined (e.g. biases, absence of metadata, the necessity to digitize etc).

A "summary table" (draft) was suggested, which brings together main characteristics of each datatype.

5.2 Discussion on data types

Paul and Janet said that there is already some work being done on getting data quality and information from old cruises. SCRIPPs and Woods Hole are working on this.

Viktor's table is a great starting point for data type information and uncertainty/metadata requirements. Catia suggested adding another row giving an idea of the possibility of getting old metadata back.

Action: Add this table to the BAMs article.

MBTs are a substantial part of the dataset and we need to tackle them. MBTS could be a possible project that can be isolated for the IQUOD project. Can we fold this into GODAR? Whatever resources are available are available to IQUOD. People are in both IQUOD and GODAR.

Action: Viktor to maintain the instrument information table and allow others to add/comment on it.

Some discussion was held on SST. SST is not a profile, and does not fit into the IQuOD mission statement. SST is not QCd by GHRSST, they use it to validate satellite data.

Action: Investigate using SST as a check for IQuOD auto QC.

A lot of MBTs have SST included and a calibration offset. MBT information is mixed and needs more understanding. Whoever looks at MBTs will have to review the historical information carefully.

Moored buoys can be removed at this stage.

Action: Lots of SST people at CLIMAR next week – talk to them about Viktor's idea of comparisons/cross validation.

5.3 Data flows, Tim Boyer

This talk describes the present data flow into the World Ocean Database (WOD) (inflow), out of the WOD to users (outflow), and aggregation with other data sources at the US NODC (confluence).

Inflow - Presently, data flows regularly to the WOD from three main near-real time sources Argo profiling floats, Pacific Marine Environmental Lab (PMEL) tropical moored buoy array, and the Global Temperature and Salinity Profile Project (GTSPP). These data sources are incorporated into WOD and made public every three months. Many other data flows, at different time steps (three monthly, yearly, irregular) are also added to the WOD. The only restriction on data entering WOD is that the data must be archived in its originally received form and freely available. For IQuOD, WOD would send out data to the manual quality control (qc) centers with automatic flags attached. When manual qc was complete, data would be sent back and uploaded into WOD with amended qc flags (and possibly amended data) for replacement. Other IQuOD related inflows would include data from original sources following IQuOD procedures, and data from other projects, such as Hydrobase, which meet or exceed IQuOD qc standards.

Outflow - Data will be stored internally to WOD with original data and amended data, original qc flags, WOD qc flags, and IQuOD qc flags. Data would then be served through WODselect where the user would be able to subset the IQuOD dataset and choose which qc flags they would like attached to the observations. A GEBICH two-tiered flagging scheme can also be implemented whereby tests passed/failed flags can be served in conjunction with a decision (bad/good) qc flag. IQuOD data would also be available as a stand-alone downloadeable flat file dataset.

Confluence - For all datasets at NODC in CF compliant netCDF format, aggregated selections of data will be prepared and served through the NODC Geoportal and THREDDS server using the definitive source for each data type. So, for instance, a user asking for all salinity data at NODC can receive all historical profile data from WOD, all near-real time profile data from GTSPP, except Argo data, which will come directly from the Argo program, and tropical moored buoy data from PMEL/NDBC. These profile data can be combined with thermosalinograph data in CF compliant netCDF from a planned NODC database (or source), coastal buoy data from NDBC, and even satellite data from Aquarius or SMOS. This system is under construction at NODC.

5.4 Discussion on data flows

Tim envisages the IQUOD QC flags as another layer over the WOD dataset. Tim has been communicating with Ruth Curry and she is happy to have the Hydrobase QC as part of IQUOD.

We aim to have standard autoQC and manual QC tests that independent people can run over data.

Questions for further consideration:

- What flags do we make available for other data types that come with temperature data?
- As we build up the IQUOD database – what flags/uncertainties do we put on data that hasn't been QCd yet?
- Do we put a DOI on the dataset? Very likely and desirable.
- Should we centralize the auto QC and the farming out of the manual QC? Probably yes.

Action: Tim to incorporate the data flow system proposed into the implementation plan.

5.5 Ocean Data Standards: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. Hernan Garcia

Hernan Garcia (Hernan.Garcia@noaa.gov), USA), Sergey Konovalov (Ukraine), Cyndy Chandler (USA), Reiner Schlitzer (Germany), Laure Devine (Canada), Gwen Moncoiffé (UK), Toru Suzuki (Japan), Alex Kozyr (USA), Greg Reed (Australia)

A universal and unambiguous quality flag (QF) scheme applicable to all data variables/parameters is recommend for use in exchange of oceanographic and meteorological data (JCOMM/IODE Ocean Data Standards and Best Practices Project, ODSBP). Published in April 2013 as a UNESCO/IOC Manuals and guides No. 54 - Volume 3: Ocean Data Standards: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. The QF scheme ensures (1) data quality consistency within a single data set and within a collection of data sets and (2) that the quality and errors of the data are apparent to the user, who has sufficient information to assess its suitability for a task (fit for purpose). The Quality Flag Scheme was developed by a committee comprising representation from several different countries. It is a two-level system. The two-level scheme enables the small number of unambiguous flags at the primary level to be optionally justified by the details represented in the second level. The primary level defines the data quality flags only (intended for data users that need only basic data quality flags). Primary-level flag values are numeric and ordered such that increasing quality flag values indicate a decreasing level of quality. The secondary level (recommended), complements the first level by providing the justification for the quality flags, based on quality control tests or data processing history, estimates of data uncertainty and errors (fit for purpose). This QF scheme has several advantages over existing ones: Small number of primary-level flag values that are numeric and ordered such that increasing quality flag values indicate decreasing level of quality. This supports identification (filtering) of data that meet a minimum quality level and assignment of quality flags to calculated parameters and facilitates data exchange and mapping between QF schemes without loss of information. The scheme was approved as recommended standard at the twenty-second Session of the IOC Committee on International Oceanographic Data and Information Exchange (IODE).

Intergovernmental Oceanographic Commission of UNESCO. 2013. Ocean Data Standards, Vol.3: Recommendation for a Quality Flag Scheme for the Exchange of Oceanographic and Marine Meteorological Data. UNESCO, Paris, France. (IOC Manuals and Guides, 54, Vol. 3.) 12 pp. (English) (IOC/2013/MG/54-3) http://www.iode.org/mg54_3

5.6 File content and data flagging, Ann Thresher

Essential metadata should be defined early to avoid expending a lot of energy later to retrieve missing information. We will need several sections for types of information that will be permanently associated with the profiles we QC. This will include instrument metadata such as XBT probe serial number, data acquisition system type and serial numbers, data acquisition software version if applicable, ship ID and callsign.

We will also need profile specific information such as latitude, longitude, date and time of the profile, and we must carry all unique ids associated with a profile, given that different groups use different systems to identify a particular profile in their databases.

All sensor data collected as part of the profile (not just T and S, Depth/Pressure) should be carried. We need to keep the raw version of data, as well as a 'best copy' version that has been through QC. We also intend to associate a QC flag and an error estimate with each value in the profile. If a calibration or conversion of the data has been done, then we need to carry this calibration information.

Finally, we need a history that records what tests were applied, what system applied the tests, what tests were failed, who applied the test, etc. We also need to define how we use the history record in more detail – how do we indicate the things we need to keep track of such as whether this profile has been through Auto QC or not and where it's failed which test.

Ann proposed that the IGOSS QC flagging system with flags 0-9 be used, as currently used by GTSPP. Finally, NetCDF 4 with the classic interface should be used so that we can more easily move forward and be backward-compatible. The Formats working group will address this in the next year.

5.7 Discussion on file content and data flagging

A metadata list will be made public to the group so we can add to it and classify each metadata type as mandatory, desirable or optional (for example). We should investigate existing metadata lists to adopt/adapt to IQUOD. Eg NODC template.

Also, we should consider adding a quality assessment to the metadata. Maybe an overall profile QC flag such as used in Argo to describe a percentage of bad data.

Ann suggested keeping history records within the database and not with data downloads.

We could consider an index file of metadata for each file in the database. What do we need in the index file? Information for QC could be useful eg, if the profile failed an auto QC test. Charles suggests including a checksum or equivalent.

Uncertainties will be assigned at the autoQC stage, and these can be adjusted during manual QC. Do we assign flags automatically if we haven't looked at the data manually? What level do they get assigned if they fail a test? This level of detail can be handled by the uncertainty and flagging task group.

Syd Levitus said that there is something to be said for transparency – people identify their own data that is incorrect.

We are missing a duplicate-checking algorithm step in the project.

Action: Ann/Bec to supply Tim with duplicate information derived from the Quota project so he can check the WOD.

Action: Simon Good to lead a working group for formats/uncertainties/flagging. Simon will start by talking with Derrick Snowden @ NOAA re using their existing formats and working with them to adapt to our requirements.

6. Specific areas of work 1

6.1 Assigning uncertainties to individual observations: experience from building HadIOD, Simon Good

HadIOD is the Met Office Hadley Centre Integrated Ocean Database (Atkinson et al. 2014, submitted to JGR). It combines sea surface temperature and profile observations and there are currently about 1.2 billion of these in the dataset. It is aimed at data assimilation so observations needed to be bias adjusted if necessary and uncertainties supplied with each. The error model used includes three uncertainty components: uncertainty in bias adjustments applied to a whole platform type, uncertainty in bias adjustments applied across an individual platform and a random uncertainty. The World Ocean Database documentation and the Abraham et al. (2013) article in Reviews of Geophysics were the basis of uncertainty estimates for the profile data. However, it was found that there is less information available about uncertainties than there is for the surface observations. Little information is given about whether quoted uncertainties are due to errors that are correlated between measurements or random, or a mixture of both. Many of the uncertainties are manufacturer specified instrument accuracies, but these may be an underestimate because for the surface observations studies into data accuracy have tended to find larger uncertainty than quoted by the manufacturers of the instruments. Therefore we need to work towards refining our uncertainty estimates. This necessarily requires collaboration. It could be useful to think about having different uncertainty estimates for different subtypes of instrument and intercomparison of data from different sources or feedback from reanalyses could help to refine estimates. There is potentially a lot that could be done and what effort that is available should be focussed on aspects that are most important to users.

6.2 Discussion on uncertainties

How do we improve on uncertainty beyond instrumental estimates? Maybe what we need is a lookup table for each data type that will simplify error assignment at least initially.

Modelers are major users and some really need uncertainties. Possibly we can use models to help us with uncertainty estimates for auto QC. We need to investigate how using models can help. What about representivity errors? How important are they? Probably they are more important are bias errors and many biases are unknown or yet to be discovered. Is it our role to provide representivity error? This type of error might be seen as the main type of error. Can we reach out to other communities to get some help from them?

Syd suggests adding satellite sea level users to the list. They might be part of the climate modelling community.

Matt: One group we haven't included yet are the ECHO consortium – they use observations to constrain their models. Uncertainties are very important to them. Models might be able to provide us with an initial guess at representivity errors. We

might be able to use models to reduce the failures in the auto QC. All models have a representivity error. How can we give a preference to one model over any others?

IQuOD needs to acknowledge that the end-users need representivity error, but we are not sure how we'll get there. We need to continue the conversation with them.

6.3 Data aggregation, Tim Boyer

This talk starts with a description (size and shape) of the World Ocean Database (WOD). It consists of 13.2 million casts (collections of profiles of different ocean variables taken at the same date/time/position). The WOD includes 12.9 million temperature profiles, 5.4 million salinity, 1 million oxygen profiles, and less than 0.5 million profiles of any other ocean variable. The time span is 1772 (Captain Cook) to present, with the majority of the data from the period 1955-present - 12 million profiles. The depth range of the data is surface to ocean bottom, with > 10 million profiles reaching to 100m, > 7 million to 400m, > 3 million to 700m, > 1 million to 2000m. The observation system for temperature has shifted from reversing thermometer/bottle (1772 - 1939), mechanical bathythermograph (MBT, 1939-1967), expendable bathythermograph (XBT, 1967-2001), to Argo (2001-present) with contributions from other instruments (thermistor chains, ship-based CTD, etc.).

Data comes to NODC and into the WOD from many sources: large scale projects such as WOCE (1990-1998), Universities, Fisheries managers, government agencies, smaller scale regional projects, data centers, institutes, and individual scientists. Data comes in many forms, from downloaded files on the internet, to CD and DVD, to older computer media such as floppy disks or reel-to-reel tape. Index cards and individual cruise reports also contain valuable material for the WOD.

There are still many data to be identified, brought or sent to NODC, and uploaded to WOD. For instance, of the four major Fram voyages, three are already in WOD, one, the Sverdrup Canadian Archipelago cruise, was recently identified as missing, and the data tracked to NCDC where it had been digitized from an old cruise report. Sippican, the primary manufacturer of XBTs, reports 7 million XBTs sold from 1962-2002, whereas the WOD contains only 2 million XBT drops. Areas of the ocean, such as the Sea of Ohkotsk, are lacking in data, but more are being added, with an increase from 30,000 casts to 43,000 casts in WOD for the area added since January, 2013.

NODC is always on the lookout for data, finding online data, talking to presenters and scientists at meetings such as AGU, keeping abreast of scientific literature, and maintaining contacts with other data centers, institutes, and individual scientists.

IQuOD participants can help by alerting NODC to data sources, or soliciting directly the data, by educating submitters in best procedures and the importance of submitting data, and persuading data holders to release without restriction their data holdings.

6.4 Discussion on data aggregation

GODAR has been funded with approximately 5K per year and this feeds to WOD. Currently NODC is not funded to continue data rescue efforts, but to collect new data. US researchers are required to provide data management plans with funding applications, and can still protect data for a time but not forever. This has made a difference to archiving data, but there are still datasets that aren't archived.

There was some discussion on pattern recognition and fuzzy logic with Igor.

7. Specific areas of work 2

7.1 Intelligent metadata, metadata recovery and bias adjustment, Simon Good

We require metadata to tell us information about data such as the type of instrument used and what processing was applied to the data. Unfortunately many profiles are missing key metadata such as the probe type and manufacturer for XBTs. This can confound attempts to derive and apply bias adjustments. We need to compile a list of what metadata we need e.g. for this kind of activity. Then we need to consider whether it is possible to recover the metadata by finding the information somewhere and adding it to the record. An alternative approach is 'intelligent metadata' where other metadata and/or features of the data are used to infer the metadata that we require. However we need to be aware that this is always going to be uncertain and we need to somehow communicate this to users.

7.2 Discussion on metadata, bias adjustments.

Many profiles have incomplete metadata, for example, many XBTs are of unknown type or manufacturer and this is critical for bias corrections. As a result, we end up guessing metadata, which coined the tern 'intelligent metadata'. The data formats group has to define the necessary metadata. Cyndy Chandler has already started determining essential metadata by polling users as to what is necessary (Cchandler@whoi.edu) (http://www.bco-dmo.org/documentation/metadata-database-design-information-and-background). We should also look at the SOOP program list of metadata.

What is intelligent metadata? Can we infer metadata from what we do have? Some adjustments have already been made to some data types, eg MBTs, but we can't tell which have been adjusted and which haven't, which will make bias adjustments difficult. We need to flag the metadata particularly if it's inferred – either/or with a quality flag and an uncertainty. A good test of intelligent metadata will be to test any scheme developed against data with known metadata.

Tim asked how do we recover metadata? We can contact people, look at cruise reports, recheck the archives for translation errors, and some is irretrievably lost. Maybe these data recovery projects can be funded in small chunks. Syd stated funding is favoring new observations, not the older ones. Crowd sourcing was discussed for metadata recovery. It is a very possible option.

7.3a Manual QC, Ann Thresher

The overall QC steps include:

- Data set assembly
- Data converted to the agreed netcdf format
- Duplicate checks completed
- Run dataset through agreed Auto QC
- Apply manual QC

We need a QC interface that allows us to apply QC flags and output the QCd profiles, and a structure to control and guide the manual QC – a way to coordinate volunteers, whether crowd sourced or scientists with free time. This will then determine how we further subset the global dataset, eg, by instrument for instrument experts, by region for regional experts. Training for manual QC is very important for consistency between operators and through time. We plan to ask IODE for assistance with this through their Ocean Teachers program.

An initial estimate of the cost for manual QC depends on many assumptions because many of the numbers are not yet well defined or known. If we need to manually QC 35% of the profiles, and an expert can QC approximately 200 profiles an hour (assuming also that most of the data to be QCd will actually contain only good data), and choosing an arbitrary cost of an expert as $30 USD, the cost will be approximately $500,000 USD to accomplish stage 1 of IQuOD. We can contain costs perhaps by making use of ‘citizen science’ or programmers who volunteer their expertise.

The cost can be reduced by ensuring that the Auto QC system is efficient at catching bad data and ignoring good data. But, we must catch MOST (if not ALL) of the bad data and this comes at a cost. We won’t know that cost until we finish designing the Auto QC system.

7.3b Crowd Sourcing for IQuOD manual quality control? Matt Palmer and Philip Brohan, Met Office Hadley Centre

Citizen science is now a well-established activity with a vast array of projects listed at www.scistarter.com. This presentation explored the potential for using "crowd sourcing" as an approach for manual quality control of IQuOD data, based on experience with the very successful www.oldweather.org.

The old weather website cost about £50k to build and requires the equivalent effort of one full-time postdoctoral researcher to maintain. In terms of digitized data, it has delivered approximately 10-15 FTE years of effort in about 2 years. A unique selling point of taking the crowd sourcing approach is the possibility of getting statistics on individual QC decisions, if the project were successful enough to entrain that level of effort.

The main challenge will be to make the QC activity rewarding and fun in some way. Old weather has a lot of historical information and depth that helps to engage citizen

scientists. It is likely that IQuOD QC would have to made into some sort of game (as has been done for e.g. digitization of Finnish newspapers). Background could be developed in terms of the different oceanic basins, including perhaps circulation and indigenous wildlife. If successful, IQuOD would be the first citizen science project to tackle the QC problem - and could blaze a trail for other variables and observational communities.

7.4a Discussion on Manual QC

We need to consider how to subset the data for the manual QC step. There are many options, eg, by date, location, data type. Volunteers and funded experts are needed to QC. We might be able to use crowd sourcing. Consensus on manual QC and training are both required. Ann suggests some half-day workshops on QC.

Ann's estimates on manual QC rates are thought to be too optimistic, the group would like to come up with a better estimate. The more we split the manual QC up, the more the costs: training, time, retraining, checking.

Action: Come up with a more realistic estimate of how many an operator can do in an hour. Chris Paver, Melissa Zweng, Francis Bringas? Put the updated costing for manual QC into the implementation plan.

Action: Need to identify experts to be involved in the manual QC workshops to pin down the QC requirements and specifications prior to the workshop. Get a half-day/day workshop up to get some manual QC methods discussed.

Ruth Curry has already committed to giving us HydroBase 3 to join the IQUOD dataset.

Action: Identify already highly QC datasets.

Some discussion was held about whether QCd data should be run through the auto QC. Ann says no, others say yes. We can't assume that the data has been QCd properly. Argo has a great QC path, but there were the Woods Hole floats that have all been greylisted. We should use the already QCd databases to test the auto QC system. Data types that will be included need to be prioritised. We can start to prioritise using Viktor's tables.

We should consider other ways to perform manual QC.

7.4b. Discussion on Crowd Sourcing

The user interface needs to be more than just the data. It will probably need a committed PI/young career scientist to lead the project. The group was very enthusiastic about the idea of crowd sourcing. A strong selling point is that it will be the first time that QC will be done this way. How can it be made rewarding and interesting? The creativity/ingenuity side of it will have to be done in collaboration with another group who have experience in citizen science. A project like this has a moderate chance of success and low risk.

The estimated effort based on the existing project is 1FTE for 3 years. How much effort is this compared to the QC needed? By using crowd sourcing, we can get statistics on the decisions made in manual QC. Can we get someone to run with a proposal? Ann suggested Ken Ridgway might be interested. If we use crowd sourcing for manual QC, another layer will be needed to check the results. Does this make the whole process inefficient – or does it save time/money?

Creation of a web app is possible and it could be made into a game (can we charge?). Can we partner with a software designer? We could also offer up digitization via crowd sourcing along with the QC as part of funding proposals.

Action: Need someone to investigate crowd sourcing for manual QC. The group should try to think of someone to take it on. We will circulate the idea to the wider IQUOD community.

Some suggested rewards for effort were: naming rights on profiles/argo floats/area of the ocean, time on the Google ship, a trip on l'Astrolabe. Possible partners for developing software might include: students, Google, app developers. Someone to produce a proof of concept would be a good start.

Action: To help the case for data recovery: take OHC for a time period, remove some data and compare the differences. Do the same for XBT biases. Use this information to include in implementation plan.

8. Refining the project structure and plan 1

8.1 and 8.2 Action items and discussion on action items

Action items were reviewed and are given in Appendix 4.

8.3 Next meeting:

Possible hosts were discussed and final location will be dependant upon funding available from the host and the need to develop interest in IQuOD in that country. Gustavo will approach the Hydrographic Office in Brazil and in Argentina. Bec will talk to Lijing Cheng about China. Catia will talk to Nico re China too.

Exact dates to be determined, but the third meeting will be around July, 2015.

9. Refining the project structure and plan 2

9.1 Funding

Steve Diggs chaired the session and led the discussion. He opened by introducing us to three google docs for editing by the group: http://goo.gl/KJ26xE. The group then discussed the mission statement in detail and made some edits.

Steve said that in the mission statement we should be very clear about what we are not going to do.

Action: Everyone to edit the google docs.

There are people out there who will translate code for you for nothing.

Action: Alison and Steve to pair to work towards the crowd sourcing idea.

There was a discussion on producing plots of data fields with and without high quality data sets. We want to put the visuals right up front to make clear the benefits of the project.

Action: Plot up a dataset gridded dataset with and without data QC. Simon has done this already with EN dataset. Is there a simulation we can do? For example, OHC with varying QC applied. Gustavo/Marlos, Simon, Bec, Tim/Melissa.

Steve suggested we come up with an elevator pitch: a simple, 2 sentence statement. Also, that we map our outcomes onto existing programs, and bend our objectives to meet the objectives of funded programs. Eg, we could focus on the MAPP program in NOAA, it is well funded. Can we map to their objectives?

Catia has done some work aligning our focus with CLIVAR, US CLIVAR etc to get endorsement. Janet said there is a meeting of US CLIVAR and we should present to them then.

Action: Steve will attend in Denver, July. Janet and Steve to coordinate.

Andreas Schiller is co-chair of GODAE and he said they can endorse IQUOD.

Action: Write to GODAE to ask for endorsement.

Peter Gleckler can endorse IQUOD via modelling community. They don't assimilate, but do evaluate their models. We could target some input to suit AR6. Runs will be beginning in 2017. We might have a first version of IQUOD ready for then.

Should we make manageable chunks for funding? Or just have a timeline with outcomes – it might make it easier to shuffle outcomes for us?

Action: If possible, members of IQuOD to see if we can get a certain number of hours to do IQUOD activities as part of your employment.

We need to quantify how much we can do on the project with the funding we have already. Then we can present this information to a funder to show what manpower we already have. Can we get a project officer? Even someone half or part time to help out?

Action: Tim, Ann, Bec and Charles to talk with Peter Pissierssens at GTSPP about maybe enlisting his help via the IODE project.

We need a 1 page summary of IQUOD to ask for an IOC/IODE recommendation. Toru and Charles can assist with IODE requirements.

9.2 Working group goals and membership

Membership and goals of the four working groups nominated during the first meeting were reviewed. A new working group (File formats, flagging and uncertainties) was nominated. Simon Good will be the group leader and some members were included. All updated group members and goals/tasks are included in Appendix 3.

9.3 IQuOD group structure review

During review of the IQuOD structure, the lack of chairs and a steering committee was identified. Co-chairs were elected (Catia Dominges and Rebecca Cowley), and a steering committee was elected (Tim Boyer, Ann Thresher, Simon Good, Matt Palmer, Susan Wijffels, Gustavo Goni, Janet Sprintall, Alison Macdonald, Toru Suzuki, Viktor Gouretski, Steve Diggs)

9.4 Further discussion

No further discussion was held at this stage.

10. Communication plans and close

10.1 and 10.2 Communication and webpage discussion

Bec introduced the efforts toward creating a web page, including a Confluence web page set up at CSIRO, a sub-page of NOAA for IQuOD and the www.iquod.org webpage. The Confluence web page was not so successful due to the labelling by CSIRO and difficulty in using it with international partners. The latter two sites just contain second workshop information at the moment. Ricardo Domingues from AOML has been working on a template based on materials Bec put together for him.

Action: Ricardo and Bec to develop the webpage at www.iquod.org and get a draft up for comment.

On the webpage, primary sponsors need to be made prominent (those that provide cash). However, everyone who gives support to the project will be acknowledged.

Action: Set up a mailing list that is useable to everyone. Make sure the steering team has a generic email address.

We need some mechanism for feedback, so we require a feedback box. User requirements surveys could also be delivered via the web page.

10.3 Other business:

SCOR funding was discussed and we need to talk with IAPSO and SCOR at least 5 months before the funding applications. Make sure what we are well prepared.

Action item: Get SCOR application in. Catia to begin the application and get back to steering team to review it.

Action: Set up some webinars during the year. Catia, Bec and steering team.

Action: Set up regular teleconferencing: Group leaders to organise based on reviewing action items and need to discuss.

Action: Set up some regular meetings between the steering team. Catia and Bec.

10.4 Review action items

Additional action items were reviewed and added to the list in Appendix 4.

10.5 Close

Gustavo and Tim thanked the attendees of IQuOD for their efforts. In return, the group acknowledged and thanked the organising committee and local organisers (NOAA) for holding the meeting.

Appendix 1. Workshop Agenda

Wednesday June 4

8.00 – 8.20	Registration and coffee		20 mins
8.20 – 8.30	**Welcome and local logistics**	NOAA representative (Tim Boyer or Gustavo Goni)	10 mins
8.30 – 8.45	**Introductions** Round the table getting people to say who they are, where they are from and main area of expertise	All	15 mins
Session 1: Setting the scene – current project structure, workplan and progress made in the last year.			
Session Chair: Viktor Gouretski. Notetaker: Simon Good			
8.45 – 9.05	**1.1 Project aims and structure** Recap of aims of the project, what happened at the first workshop and overview of the current project structure, action items from last meeting, outcomes from Ocean ScienceTown Hall. Corporate Image.	Catia Domingues,	20 mins
9.15 – 9.35	**1.2 Recap on goals for Auto QC, Manual QC, Data aggregation and Data assembly groups** – recap goals from first workshop.	Ann Thresher,	20 mins
9.35 – 9.50	**1.3 The scientific implementation plan and plans for CLIVAR endorsement** Discuss the scientific implementation plan outline, the importance of CLIVAR endorsement etc.	Matt Palmer	15 mins
9.50 – 10.10	**1.4 Discussion** - after a year to reflect, did we do anything wrong, is anything missing from the current project structure, what has gone well etc.	All	20 mins
10.10 – 10.30	Tea, coffee, biscuits		20 mins
Session 2: Auto QC group benchmarking results			
Session Chair: Catia Domingues. Notetaker: Matt Palmer			
10.30 – 10.40	**2.1 Auto QC group** – brief description of current work plan and the progress over the last year.	Bec Cowley	10 mins
10.40 – 10.50	**2.2 Description of AOML tests**	Francis Bringas	10 mins

10.50 – 11.00	**2.3 Description of NODC tests**	Tim Boyer	10 mins
11.00 – 11.10	**2.4 Description of CSIRO tests**	Bec Cowley	10 mins
11.10 – 11.20	**2.5 Description of Met Office tests**	Simon Good	10 mins
11.20 – 11.30	**2.6 Description of ICDC tests**	Viktor Gouretski	10 mins
11.30 – 11.50	**2.7 Auto QC results** The auto QC benchmarking tests/results. Can we quantify the cost for the Manual QC step? Are there any other groups that should be involved?	Ann Thresher, Viktor Gouretski	20 mins
11.50 – 12.30	**2.8 Discussion of auto QC results** Discussion of auto QC results / should there be publications / where to go from here. Can we converge and agree on autoQC tests?	All	40 mins
12.30 – 13.30	Lunch (self-funded)		60 mins
Session 3: Invited talks from data users and producers			
Session Chair: Tim Boyer. Notetaker: Alison Macdonald			
13.30 – 13.50	**3.1 Invited user talk 1**: Operational SI Climate Forecasting Data Requirements: Distribution, Types, QC	David Behringer	20 mins inc. questions
13.50 – 14.10	**3.2 Invited talk 2**: Current status of the Japanese XBT data reconstruction and related ocean-atmosphere analysis work.	Masayoshi Ishii	20 mins inc. questions
14.10 – 14.30	**3.3 Invited user talk 3**: The importance of QC'd data for models/reananlysis.	Jim Carton	20 mins inc. questions
14.30 – 14.50	**3.4 Invited talk 4**: The QC of seal tag data.	Fabien Roquet	20 mins inc. questions
14.50 – 15.20	Tea, coffee, biscuits		30 mins
15.20 – 15.40	**3.5 Invited talk 5**: QC methods at CORIOLIS/IFREMER	Christine Coatanoan	20 mins inc. questions
15.40 – 16.00	**3.6 Invited user talk 6**: How might IQuOD interface with climate Model Intercomparison Projects (MIPs) and obs4MIPs?	Paul Durack for Peter Glecker	20 mins inc. questions
16.00 – 16.30	**Catch up time if needed**		30 mins

Session 4: User requirements and interactions			
Session Chair: Ann Thresher. Notetaker: Bec Cowley			
16.30 – 16.50	**4.1 User requirements gathering** – experience from the ESA Climate Change Initiative project on SST	Simon Good	20 mins inc. questions
16.50 – 17.30	**4.2 Discussion on user requirements** What do the user requirements tell us about IQUOD? Do we need to do more user requirement gathering? Should we focus in on satisfying particular user group(s)?	All	40 mins
18.30 onwards	Dinner location to be advised (self-funded)		

Thursday June 5

8.00 – 8.20	Tea, coffee		20 mins
Session 5: Data types and flows			
Session Chair: Bec Cowley. Notetaker: Catia Domingues			
8.20 – 8.40	**5.1 Data types** Review the types of data (different instruments, date ranges etc.) in the historical record and recap decisions made at last workshop about which we are going to QC.	Viktor Gouretski	20 mins
8.40 – 8.55	**5.2 Discussion on data types** Discussion to reaffirm/change the types of data we are handling, discuss the implications of including / not including particular types.	All	15 mins
8.55 – 9.15	**5.3 Data flows** Description of how the end to end processing chain will/might work for IQuOD (e.g. original data into WOD -> Auto QC locally at NODC or somewhere else? -> WOD -> Manual QC group -> back to WOD, will there be data releases held elsewhere and if so how do we combine IQuOD data with other data e.g. Argo for those who want all data?	Tim Boyer	20 mins
9.15 – 9.30	**5.4 Discussion on data flows**	All	15 mins
9.30 – 9.50	**5.5 ISO endorsed flagging system.**	Hernan Garcia	20 mins
9.50 – 10.10	**5.6 File content and data flagging** What do we need in the files to make the data useful and easy to use?	Ann Thresher	20 mins
10.10 – 10.35	**5.7 Discussion on file content and data flagging**	All	25 mins
10.35 – 10.55	Tea, coffee, biscuits		20 mins
Session 6: Specific areas of work 1			
Session Chair: Matt Palmer. Notetaker: Ann Thresher			
10.55 – 11.15	**6.1 Assigning uncertainties to individual observations –** experience from building HadIOD	Chris Atkinson, Met Office (presented by Simon Good)	20 mins
11.15 – 11.35	**6.2 Discussion on assigning uncertainties** – how to go about assigning uncertainties	All	20 mins

11.35 – 11.50	**6.3 Data aggregation** – getting data into WOD that aren't already there	Tim Boyer	15 mins
11.50 – 12.10	**6.4 Discussion about data aggregation**	All	20 mins
12.10 – 13.10	Lunch (self-funded)		60 mins
Session 7: Specific areas of work 2			
Session Chair: Gustavo Goni. Notetaker: Ann Thresher			
13.10 – 13.25	**7.1 Intelligent metadata, metadata recovery and bias adjustment –** How we might approach this and how it would relate to XBT bias adjustment work. Can we quantify the cost of metadata recovery efforts? What techniques can we use in place of missing metadata?	Simon Good	15 mins
13.25 – 13.45	**7.2 Discussion about intelligent metadata**	All	20 mins
13.45 – 14.15	**7.3 Manual QC** – existing methods / software / need for training to use them etc. Can we quantify the cost of ManualQC? How do we tackle the problem – regional experts, instrument type, time period? Crowd sourcing?	Ann Thresher, [Matt Palmer (crowd sourcing)]	30 mins
14.15 – 15.15	**7.4 Discussion on how to approach the manual QC**	All	60 mins
15.15 – 15.45	Coffee break, workshop photo		30 mins
Session 8: Refining the project structure and plan 1			
Session Chair: Simon Good. Notetaker: TBC			
15.45 – 16.30	**8.1 Action Items**	All	45 mins
16.30 – 17.30	**8.3 Discussion on action items, working group goals and membership** 1) What should the task groups achieve in the next year (or beyond) and by when? 2) What input is needed from other groups and by when?	All	60 mins
17.30 – 17.40	**10.4 Agree date/location of next meeting**		10 mins
18.30	Dinner at location to be confirmed (self funded)		

Friday June 6

8.00 – 8.20	Tea, coffee	20 mins
Session 9: Refining the project structure and plan 1		

Session Chair: Steve Diggs. Notetaker: Bec Cowley			
8.20 – 8.50	**9.1 Funding** – what difficulties are there with people to have funding to work on IQuOD, what opportunities are there to get funding? If suitable opportunities are available, get volunteers to write the proposals. How should we structure the funding for IQUOD (eg – look at GO-Ship funding). IQuOD has global coordination with regional funding.	Steve Diggs to lead discussion	30 mins
8.50 – 9.05	**9.2 Working group goals and membership** – Finish the Manual QC group.	All	15 mins
9.05 – 9.25	**9.3 IQUOD group structure.** Steering team, project support, executive team?	All	20 mins
9.25 – 9.55	**9.4 Further discussion** – confirm what key outputs (papers / technical documents / software / data) we want to achieve in the next year.	All	30 mins
9.55 – 10.15	Tea, coffee		20 mins
Session 10: Communication plans and close			
Session Chair: Catia Domingues Bec Cowley – to lead discussions. ***Notetaker: Ann Thresher***			
10.15 – 10.45	**10.1 Discussion on internal communication plans** (how should we communicate within the project, introduction to website plans)	Bec Cowley – website intro.	30 mins
10.45 – 11.15	**10.2 Discussion on external communication** – Communication and outreach over the next year, including communication with data generators?	All	30 mins
11.15 – 11.35	**10.3 Any other business** that has arisen during the meeting		20 mins
11.35 – 11.55	**10.4 Agree on final list of actions from the meeting**		20 mins
11.55 – 12.05	**10.5 Final words and close**.		10 mins

Appendix 2. Participant list.

Name	Affiliation
Catia Domingues	ACE-CRC
Bec Cowley	CSIRO Marine and Atmospheric Research
Ann Thresher	CSIRO Marine and Atmospheric Research
Tim Boyer	NODC/NOAA
Gustavo Goni	AOML
Charles Sun	NODC/NOAA
Francis Bringas	AOML
Alison Macdonald	Woods Hole Oceanographic Institution
Matt Palmer	Met Office Hadley Centre
Simon Good	Met Office Hadley Centre
Viktor Gouretski	University of Hamburg
Hannah Dean	Interagency Ocean Observation Committee, Consortium for Ocean Leadership
Igor Belkin	University of Rhode Island
Christine Coatanoan	IFREMER
Jim Carton	Atmospheric and Oceanic Science Center/University of Maryland
Syd Levitus	Atmospheric and Oceanic Science Center/University of Maryland
Ariel Hernan Troisi	Servicio de Hidrografia
Molly Baringer	AOML
Masayoshi Ishii	Meteorological Research Institute
David Behringer	NOAA/NCEP/Environmental Modeling Center
Toru Suzuki	Marine Information Research Center
Melissa Zweng	NODC/NOAA
Igor Smolyar	NODC/NOAA
Olga Baranova	NODC/NOAA
Krisa Arzayus	NODC/NOAA
Ricardo Locarnini	NODC/NOAA
Paul Durack	Lawrence Livermore National Laboratory, US
Alex Kosyr	CDIAC
Steve Diggs	UCSD
Janet Sprintall	UCSD
John Antonov	UCAR Project Scientist at NODC
Chris Paver	National Oceanographic Data Center
David Legler	Climate Program Office-NOAA, US
Steve Piotrowicz	Climate Program Office-NOAA, US
Hernan Garcia	NOAA/NESDIS/NODC
Dan Seidov	NOAA NODC/MDSD OC1
Alexey Mishonov	ESSIC/CICS-MD, University of Maryland & Ocean Climate

	Lab/NODC/NESDIS/NOAA Affiliate
Candyce Clark	CPO
Joel Levy	CPO
Marlos Goes	AOML

Appendix 3. Working group membership and tasks 2014-2015.

Data formats, uncertainty, flags

Led by Simon Good.

Members: Derrick Snowden, Ann Thresher, Paul Durack, Steve Diggs, Toru Suzuki, Charles Sun, Simon Wotherspoon?, Hernan Garcia?, Chris Paver?

Goals (finish date):

- Gather a group to help. (August, 2014)
- Establish a baseline of what formats and metadata is available already. (September, 2014).
- Investigate algorithms already available for intelligent metadata. (September, 2014).
- Get started on uncertainties. Viktor, Simon. (September, 2014)
- Meet and review to discuss in September
- Decide on formats, flags, metadata. (June 2015.)

Aggregation working group:

Led by Catia Domingues.

Members:

Tim Boyer, Gustavo Goni, Viktor Gouretski, Nathan Bindoff, Steve Diggs, Marty Hidas, Sebastien Mancini, Guy Williams, Uday Bhaskar, Toru Suzuki, Sergey Gladyshev?, Molly Baringer?

Goals:

- Provide an inventory of the details of the problems for available historical data/metadata
- Identify and contact PIs with data. Lobby at high levels. Report back to next meeting.
- Begin acquisition of readily available sources of QCed/flagged data into WOD.
- Begin engaging parties in acquisition. Report back at next workshop.
- Investigate crowd sourcing options wrt digitizing or finding out about more data that is available. Crowd sourcing metadata. Report back at next workshop.

Auto QC working group:

Led by Rebecca Cowley.

Members:

Ann Thresher, Tim Boyer, Jeff Dunn, Shoichi Kizu, Gustavo Goni, Viktor Gouretski, Guy Williams, Matt Palmer, Simon Good, Francis Bringas, Paul Durack, Fabien Roquet, Christine Coatanoan

Goals:

- Finalize the auto QC assessments. December, 2014
    - Investigate Objective Analysis techniques used by IFREMER/CORIOLIS. Contact Uday to see what he has done with this algorithm.
    - Refine the formats used for testing.
    - Refine the databases used for testing.
    - Use a reduced profile dataset to allow us to tune our QC methods and then compare results.
    - All participants in testing to write a list of the tests and how they are applied.
    - Decide on the auto QC tests and steps to be applied.
- Use the whole WOD to also compare every test system and produce some statistics. Viktor, Ann. December 2014
- Write a report/paper on the outcomes of the testing. June 2015
- Pass on algorithms to Tim for implementation. June 2015

Goals (longer term):

- Gather together a group of experts in individual instrument types to develop a scheme for assigning error estimates (Viktor)
- Provide a clear statement of what the AutoQC group would like to achieve
- Document a standard list of tests

Manual QC working group:

Led by Ann Thresher

Members:

Shoichi Kizu, Matt Palmer, Francis Bringas, Simon Good, Uday Bhaskar, Ping Robinson, Alison Macdonald, Rebecca Cowley, Lisa Lehman, Giles Reverdin?, Alejandro Orsi?, Guy Williams, Toshio Suga?, Melissa Zweng

Goals:

- Hold teleconferencing/communications via Google docs, and organise and hold a manual QC workshop to:
    - Define codes and flags and tests to be applied. Some of these will be instrument specific, others will be general.
    - Establish rules for quality code use, decide whether we can change data.
    - Document codes and procedures and provide detailed instructions for QC. **June, 2015**
- Establish a group of experts to advise on regional issues or issues specific to a particular instrument type. **Viktor, already on task list.**
- Provide an estimate of costs of doing manual QC for funding applications. **Already on task list.**

- Investigate Crowd Sourcing. ??
- Agree on software tools and get it written. Final software is dependant on file formats. Should be platform independent and portable. Need to source someone to write it.

Goals (long term):

Set up a group to monitor and assess the global manual QC as it is returned to the master database.

GDAC working group:

Led by Tim Boyer

Members:

Nathan Bindoff, Steve Diggs, Catia Domingues, Thierry Carval?, Simon Good, Susan Wijffels

Goals:

- Investigate possibilities for GDAC locations and define GDAC roles and resources needed (NODC, Coriolis, ESGF, Met Office)
- Define & track metrics (over different steps workflow/final product). Usage, who is using the GDAC?
    - Use the stats for WOD/CCHDO/Argo in the implementation plan (due to fit when plan is due).

Appendix 4. Action Items.

No.	Item	Who	When
1	Scientific Implementation plan V0.1. Incorporate a timeline. Get feedback from IQuOD members during development. Get feedback from community on version 1.0. Incorporate Simon/Viktors tables, Tim's data flow and Manual QC costing into the plan. Also plots of OHC with and without QC'd data, or with/without XBT biases or with /without selected data.	Matt Palmer, Catia Domingues & task group leaders	June 30, 2014. Final (1.0) version December, 2014
2	Make a summary of lessons learned from SST, ICOADS, GTSPP, GOSUD, etc efforts. Talk to Peter Thorne (and about crowd sourcing). Talk to Kate Willett.	Catia Domingues, Bec Cowley, Simon Good, Tim Boyer	Next week, June 30
3	Produce a workshop report	Bec Cowley	August 1, 2014
4	Write a workshop report summary in EOS. Use this in funding applications as required.	Bec Cowley, Matt Palmer, Paul Durack, Catia Domingues, Janet Sprintall (for advice)	August 1, 2014
5	Write a scientific review article in BAMS. Based on Scientific Implementation plan. Use this in funding applications as required.	Bec Cowley, Paul Durack, Matt Palmer, Catia Domingues, Viktor Gouretski, anyone else?	December, 2014 (tie in with SIP)
6	Auto QC group to continue on automated test comparison. Refine the testing requirements, produce a paper on the outcomes. Investigate changing the size of the test datasets. Talk with ocean modellers about lessons learnt wrt QC. Investigate using salinity & synthetic salinity QC	Auto QC group (Bec Cowley, Ann Thresher, Simon Good, Viktor Gouretski, Toru Suzuki, Francis Bringas, Tim Boyer)	June, 2015

No.	Item	Who	When
	techniques in auto QC process. Investigate using SST as a check for IQuOD QC and cross validation. Refer to task list in section 2.8.		
7	Create a list of regional/instrumental experts who will be willing to contribute to the Manual QC aspect of the project.	Viktor	August, 2014
8	Start a list of users and user requirements, encourage user requirement feedback.	Simon Good	August, 2014
9	Group leaders to maintain regular meetings between workshops. Conference calls.	Catia Domingues, Bec Cowley, Tim Boyer, Ann Thresher and Simon Good	Twice yearly, more often
10	Establish the uncertainty and formats task team. Set goals and membership.	Simon Good	August, 2014
11	Contact Korean manual QC expert	Tim Boyer	June 30, 2014
12	Send Tim Boyer a list of duplicates identified in WOD during the Quota project.	Bec or Ann	December, 2014
13	Try and formalize the manual QC procedure and relate it to efficient auto QC procedures. First draft for manual QC workshop.	Ann Thresher	June 2015.
14	Organise a manual QC workshop in conjunction with IQUOD workshop 3. Identify experts to be involved and pin down QC requirements prior to the workshop.	Ann Thresher	June 2015
15	Investigate if clones of IQUOD dataset might be held at Met Office, Coriolis, ESGF/ESCOG, cloud servers.	Simon Good, Matt Palmer	Done

No.	Item	Who	When
16	Investigate how 2-way communication/feedback between users and IQUOD might work (using ESGF/ESCOG)	Matt Palmer, Catia Domingues	June 2015
17	Talk with WOD/Jim Potemra about Quota-style QC'd Pacific data for auto QC benchmarking	Bec/Catia	August, 2014
18	Viktor to maintain a table of data types (as shown in his presentation) that will be incorporated into IQUOD, their priorities etc, and allow others to contribute to it. Put it on the website?	Viktor	Ongoing.
19	Incorporate the data flow system proposed by Tim Boyer into the SIP.	Tim	In time for SIP
20	Assess the effect of reduced metadata on the quality of analysis.	Marlos Goes (volunteered by Gustavo)	In time for SIP.
21	Come up with a more realistic number of profiles per hour for manual QC	Ann, Chris Paver, Melissa Zweng, Francis	In time for SIP
22	Identify already highly QCd databases and start collecting them	Tim (collection), everyone else for input & information	Ongoing
23	Approach China, Argentina, Rio to host next year's meeting	Bec, Catia, Gustavo	June 30, 2014
24	Write to GODAE to ask for endorsement	Catia	August, 2014
25	Write to Peter Gleckler (WCRP/DOE) to ask for endorsement.	Paul	August, 2014
26	Steve will attend US CLIVAR in Denver	Janet and Steve	July, 2014
27	Plot up a dataset gridded dataset with and without data QC. Aim is to show that the QC of data and presence of metadata is important.	Gustavo/Marlos, Simon, Bec, Tim/Melissa.	July 1
28	Everyone to edit the google docs at http://goo.gl/KJ26xE	All	ongoing

No.	Item	Who	When
29	When anyone becomes aware of funding opportunities let everyone know.	All	ongoing
30	Set of slides for US CLIVAR to Steve.	Catia with input from others as required.	July 1
31	Investigate crowd sourcing for funding.	Alison and Steve	December, 2014?
32	Make a representation to IODE in March, 2015 to ask for assistance with project support.	Toru to talk with Peter Pissierssens next week and get back to us. Work with Charles.	Ongoing till March 2015.
33	Get a draft of the webpage up for comment.	Ricardo Domingues and Bec. Olga Baranova able to assist.	Next 2 weeks.
34	Set up a mailing list that is useable to everyone. Make sure the steering team has a generic email address.	Bec, Ricardo, Olga.	With website.
35	Set up some webinars during the year. Webex?	Catia and Bec and steering team	June, 2015
36	Set up regular outcome related or milestone related teleconferencing.	Group leaders	based on reviewing action items and need to discuss
37	Set up some regular meetings between the steering team	Catia and Bec	?
38	Keep track of upcoming conferences for opportunities to promote IQuOD. Use the science report and meeting report as a basis for these talks.	All	June, 2015
39	See if we can assign a set amount of time to IQUOD. Then we can present this information to a funder to show what manpower we already have.	Everyone	August 2014
40	Investigate crowd sourcing for manual QC	Volunteer from wider community	June, 2015
41	Start the application process for SCOR funding	Catia	SCOR applications due March, 2015