

# Surface Ocean CO<sub>2</sub> Atlas (SOCAT) Automation Workshop



Seattle, WA, USA, October 21-23 2014



SOCAT Automation Participants (left to right):  
Camilla S. Landa, University of Bergen, Bergen, Norway  
Are Olsen, University of Bergen, Bergen, Norway  
Ansley Manke, NOAA/PMEL, Seattle, WA  
Steve Jones, University of East Anglia, Norwich, UK  
Eugene Burger, NOAA/PMEL, Seattle, WA  
Karl Smith, UW/JISAO/NOAA, Seattle, WA  
Benjamin Pfeil, University of Bergen, Bergen, Norway  
Biva Shrestha, Oak Ridge National Laboratory, Knoxville, TN  
Kevin O'Brien, UW/JISAO/NOAA, Seattle, WA  
Denis Pierrot, NOAA/AOML, Miami, FL

Workshop support provided by:



# Surface Ocean CO<sub>2</sub> Atlas (SOCAT) Automation Workshop

## Contents

1. Introduction and Workshop goals
2. Day 1
  - 2.1. Status of SOCAT v3 QC
  - 2.2. Automation Dashboard walkthrough
  - 2.3. Lessons learned from SOCAT v3 data ingestion using automation dashboard
  - 2.4. Ocean Metadata Editor demonstration and discussion
  - 2.5. OME connection to automation dashboard and workflow
  - 2.6. Discussion of DOI's
  - 2.7. Second level QC capabilities to add to SOCAT v4 automation
3. Day 2
  - 3.1. OEM/Dashboard working group
  - 3.2. 2nd level QC working group
  - 3.3. Crossover discussion
4. Day 3
  - 4.1. SOCAT roadmap for version 3,4,5
  - 4.2. Integration of 2nd level QC into SOCAT workflow
  - 4.3. Discussion of paper on SOCAT data management
  - 4.4. Technical discussions on communication between the Sanity Checker and the Dashboard
5. Meeting Adjourned

## 1. Introduction and Workshop goals

Kevin opened up the workshop, welcoming the participants and explained the logistics of meeting at NOAA's Pacific Marine Environmental Laboratory in damp Seattle, Washington.

The primary goals of the workshop were to:

- Confirm deadlines and time schedules for SOCAT v3 quality control.
- Define deadlines for SOCAT v4 data ingestion and quality control.
- Facilitate hands-on workshop sessions to finalize steps required in order to open data ingestion dashboard for public use in SOCAT v4.
- Facilitate hands on workshop sessions to integrate additional level second level QC routines into the automated data ingestion workflow for SOCAT v4.

## 2. Day 1 Session

Day 1 of the workshop was designed to bring all technical team participants up to speed with what has been happening with the automated data management framework, and to clearly define the tasks remaining before the dashboard is ready for general use by the SOCAT community. This day also provided an opportunity to introduce the SOCAT v3 QC system to those unfamiliar with it. It was also on this day the working groups and working groups tasks were defined for day 2.

### 2.1 Status of SOCAT v3 QC

Kevin and Karl report that SOCAT v3 QC is officially open for the start of QC. QC can begin immediately for SOCAT v3 data.

SOCAT v3 QC will wrap up sometime in mid January, 2015, and the release of SOCAT v3 is scheduled for Sept, 2015 to coincide with the SOLAS Open Science Conference in Kiel.

It is noted that there will be a SOCAT v3 QC webinar scheduled for day 2 of the meeting (October 22) at 8am PST that will be led by Kevin and Cathy Cosca.

## **2.2 Automation Dashboard walkthrough (Kevin)**

As there were attendees who hadn't seen the automation dashboard demonstrations at the SOCAT workshop in Bergen, Kevin provided a walkthrough of the dashboard and it's functions.

Are asked about the v4 resubmission process for a cruise that has been suspended. Karl pointed out that in order to resubmit a cruise, it is necessary to suspend it first. Are wonders how this will work when QC is shutdown. We point out that the automation system will mean that the QC system is always running so that QC and data ingestion can be done simultaneously, and the automation system will keep track of which version of SOCAT the new or resubmitted data should be associated with.

Steve Jones asks what we should do in the situation where QC is finished, but the the official release has not happened, and an issue is found with a cruise. Should the PI be able to suspend it and resubmit? Are points out that if a DOI has already been issued, than the data cannot be modified, and therefore cannot be suspended and fixed for that version of SOCAT. The issues and flaws should be noted in the release notes for that particular SOCAT. Are points out this was done for SOCAT v2.

## **2.3 Lessons learned from SOCAT v3 data ingestion using automation dashboard (Camilla, Benjamin)**

Camilla gave a very informative presentation on some of the issues that she and Benjamin had experienced using the automation dashboard for ingesting v3 cruise data into the SOCAT QC system. It is worth noting that all v3 data was fed through the automation system, thereby giving us a chance to work out issues without burdening scientists with an untested system.

Camilla mentions that overall, the system was relatively easy to learn and use. However, she did have some issues, which are listed below:

- Cryptic error messages were confusing
  - a. It was agreed by all that in general all of the error messages from throughout the system need to much more clearly explain the conditions that caused the error
- Mooring data had identical Expocodes if they had the same national code and start date
  - a. Moorings and new platform types will provide a challenge for v4 and beyond.
  - b. Moorings and new platform types will likely need a different set of metadata entered than cruises - which will require correct OME templates
  - c. Expocodes will also be different for moorings and other platforms
    - Benjamin is participating with other groups (GOA-ON) on the best way to assign expocodes for these non-ship platforms.
      - Benjamin will report back to the SOCAT automation team with updates in this area.
    - Kevin suggests that WMO ID's might be a way to have a unique identifier in SOCAT
    - Are suggests that a mooring expocode could look like: MOOR*WMOId*YYYYMMDD, where *WMOId* is the assigned WMO ID for the platform
- Identical columns in the data files

- a. All were in agreement that identical columns in a file should mean the file is rejected and needs to be fixed by PI
- b. Quality flag columns need to have unique header names, i.e. "SST\_QF" instead of just "Quality Flag" or "QF"
- Lots of duplicate date/time warnings
  - a. These occur when there are not seconds in the data files
  - b. Are comments that we had decided that seconds are required for data submission.
  - c. Benjamin argues that we should allow the PI to approve automatically adding artificial/synthesized seconds
    - This will allow many files to pass through the sanity checker by avoiding time conflicts
    - Are wants to make sure that if this happens, this information is written into the header of downloaded files.
    - Karl describes current (v3) behavior: if seconds are not provided they are set to zero with the approval of the data submitter. This is just to relieve the submitter of the burden of adding seconds when they are not significant. No attempt is made to resolve duplicate times with non-zero seconds.
- Camilla offers the following suggestions based upon her experience:
  - a. Add a short summary of sanity checker warnings on top of list of errors
    - Steve is working on this already
  - b. Use "platform name" rather than "ship name" to avoid confusion with other platform types
  - c. Add xCO2 atm wet to allowed parameters in data column checker
  - d. All data should be entered by PI as UTC/GMT
    - if Not, we risk problems with daylight savings times, etc.
  - e. Would be useful to have better capabilities to select individual cruises by categories - i.e., submitted, not submitted, etc.

## 2.4 Ocean Metadata Editor demonstration and discussion (Biva)

Biva presents the SOCAT custom OME for discussion. And there's lots of discussion. Here is the list of changes that Biva will be making to OME following the discussion:

1. Need versioning of metadata at dashboard level
2. Variable Standard list but still free form? Please provide the standard list to Biva.
3. Confirm the NODC vessel code before going to the database?
4. Separate units of the variable and description of the variable into separate fields.
5. Additional equilibrium information below Equilibrator Vented
6. Water\_Flow\_Rate (L/min)\*: move to Sampling and Equilibrator Design section and rename equilibrator water flow(L/min)
7. Rename Atmospheric Pressure (sea level) to just Atmospheric Pressure
8. Sampling and Equilibrator Design: just below the Atmospheric Pressure
9. Surface Water CO2 Method Description should have a system manufacturer description at the very top
10. Other Sensors: add type of sensor
11. remove Append Measurement or Calibration reports here:
12. remove Data Set Link:
13. Add save as draft option
14. remove Geographical Region:
  - a. make bounds read only (will be generated from data file)
15. Accuracy SST: numeric and add degree celsius
16. Precision SST: ditto as 17
17. Atmospheric Pressure (sea Level): add (not equilibrator pressure)

18. Sampling and Equilibrator design: dropdown with other
19. Co2 sensors, measurement method Co2: dropdown with others
20. Calibration of Co2 gases: convert to 2 questions 1) traceability drop down with others 2) Number of non-zero gas standards(numeric)
21. Accuracy Tequ: ditto as 17
22. Precision Tequ: ditto as 17
23. Accuracy Pequ: numeric with hPa as unit
24. Precision Pequ: ditto as 25

## 2.5 OME connection to automation dashboard and workflow (Steve J)

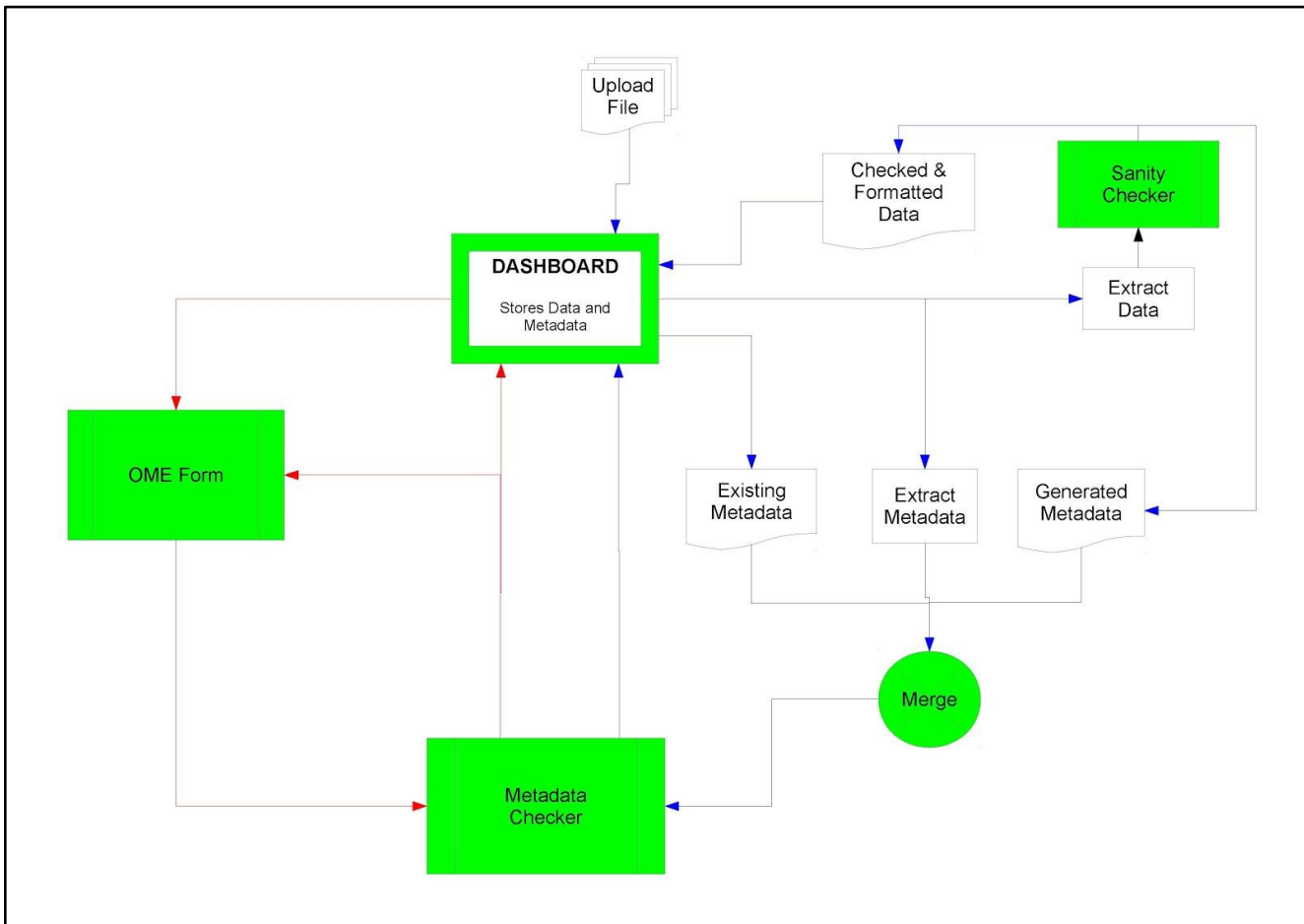


Figure 1. SOCAT v4 Dashboard and Metadata workflow

- Steve Jones provides an overview of the workflow of metadata ingestion using the automation dashboard (Figure 1).
- The metadata checker for OME metadata still needs to be written, and Steve will be working with Biva to link the processes together.
- It will be possible to edit the metadata for several cruises at once. Likely will be done by cycling through pre-filled OME forms and allowing user to edit the bits and pieces that differ between cruises (i.e., expocodes, etc)
- Start and end dates will not be entered by user in OME, but rather generated from actual data in file.
  - Geographical limits will also be generated from data file

## 2.6 Discussion of DOI's (ALL)

- It is generally agreed by all that when submitting data to CDIAC, we will expect some kind of “acceptance” response from Alex. This is to
  - a. ensure that the data was received and archived
  - b. generate DOI's for original data
    - these DOI's will then migrate into the proper SOCAT data files
- It will be up to individual PI's that do not submit their data to CDIAC to inform SOCAT of any DOI's generated for original data. SOCAT will not actively seek these DOI's out.

### 2.7 Second level QC capabilities to add to SOCAT v4 automation (Are, Denis)

- Are presently the matlab scripts which generate graphics to show data outliers
- We marvel at the informative plots Are and Denis show, and agree we can add them to the dashboard preview/QC system.
- Details of which to add will be discussed during the day 2 workshops

## 3. Day 2 Activities

Day 2 is a hands on workshop day. There will be two workshops:

1. OME/Dashboard integration workshop (Karl, Biva, Steve)
2. 2nd level QC functions workshop (Are, Denis, Ansley, Camilla)

Also on Day 2, to open the day, Cathy Cosca and Kevin O'Brien host a SOCAT v3 QC Webinar to introduce the QC system to SOCAT scientists. Unfortunately, there was a large group of international scientists who either weren't aware they had to, or couldn't, call in to the webinar. Those unfortunate folks were limited to a video only presentation of the webinar. However, the webinar was recorded, and uploaded to YouTube before the end of the day so that folks could view and hear the complete webinar. As of the writing of this document, that video has received 73 views and one thumbs up.

### 3.1 OEM/Dashboard working group (Karl, Steve, Biva)

#### Working group report written by Karl Smith:

- Biva, Karl and Steve get SOCAT OME running at PMEL using Java 7
  - Karl will check updated code into SOCAT Google code repository so it can be accessed by others if necessary
  - Biva is now also a co-owner of the code repository.
- Karl will be able to use the SOCAT OME directly in the Eclipse IDE for easy debugging and testing
- The team was able to configure the SOCAT OME required SQL database for use as well.
  - Biva provided scripts to generate the database tables required by SOCAT OME
    - This will also be included in SOCAT Google code repository
  - Currently, SOCAT doesn't need all of the authentication that SOCAT OME can provide, since it is behind a firewall and a VPN. If it is ever moved out of the VPN, and write access is needed, stricter authentication can be added.
- The team was successful in loading an XML File using “file://” rather than “http://”.
  - Currently the system will allow folks to store drafts on their local computer for editing later.
  - SOCAT OME will likely need modifications to pass XML back and forth as a Java Object in order to work with the automation system and within the VPN. Karl, Biva and Steve will determine what needs to be done to accomplish this.

- The team discussed how OME would need to save metadata files and drafts using a prepended Expocode to ensure that the directory structure on the filesystem matches with the functionality of the dashboard and the data files.
- Biva also instructed Karl and Steve what would have to be modified in OME in order to add fields to the metadata form.

### 3.2 2nd level QC working group (Are, Ansley, Denis, Camilla)

Are and Denis worked with Ansley and Camilla to show the 2nd level QC routines that we would like to add to the automated data ingestion workflow. These routines will be translated from matlab scripts to Ferret scripts, to more easily fit into the LAS workflow, but also to remove any licensing requirements. The following plots will be added to the LAS QC preview system so data submitters can check their data prior to submission to SOCAT QC system:

1. Position map
2. Thumbnails (see figure 2):
  - Latitude and Longitude as a function of time (to locate in time any outliers)
  - Ship speed as a function of time
  - Pressures: Pressure\_atm, Pressure\_equilibrator, NCEP pressures, vs time
    - i. Are suggests we may want to include difference of Pressure\_atm and Pressuer\_equil on a secondary y-axis
  - Temperatures: Temperature, Temperature\_equilibrator, Difference, vs time
  - XCO2's: XCO2\_water\_equi\_temp\_dry\_ppm, xCO2\_water\_equi\_temp\_wet\_ppm, xCO2\_water\_sst\_dry\_ppm, xCO2\_water\_sst\_wet\_ppm, vs time
  - Salinity: sal, woa\_sss vs time
  - fCO2\_recommended, vs time
  - fCO2 vs salinity
  - fCO2 vs intake temperature
3. Sources of recomputed fCO2
4. Comparison with other cruises:
  - 4 plots: fCO2\_rec vs temp, and fCO2\_rec vs salinity + fCO2\_rec vs lat and vs lon
  - Are suggests two sets of these are made: one for open ocean and one for coastal
5. Time series plots: fCO2, temp, sal, lon, lat on 5 lines, same time axis
6. Monthly plots showing cruise with all-data and all-data adjusted for trend.

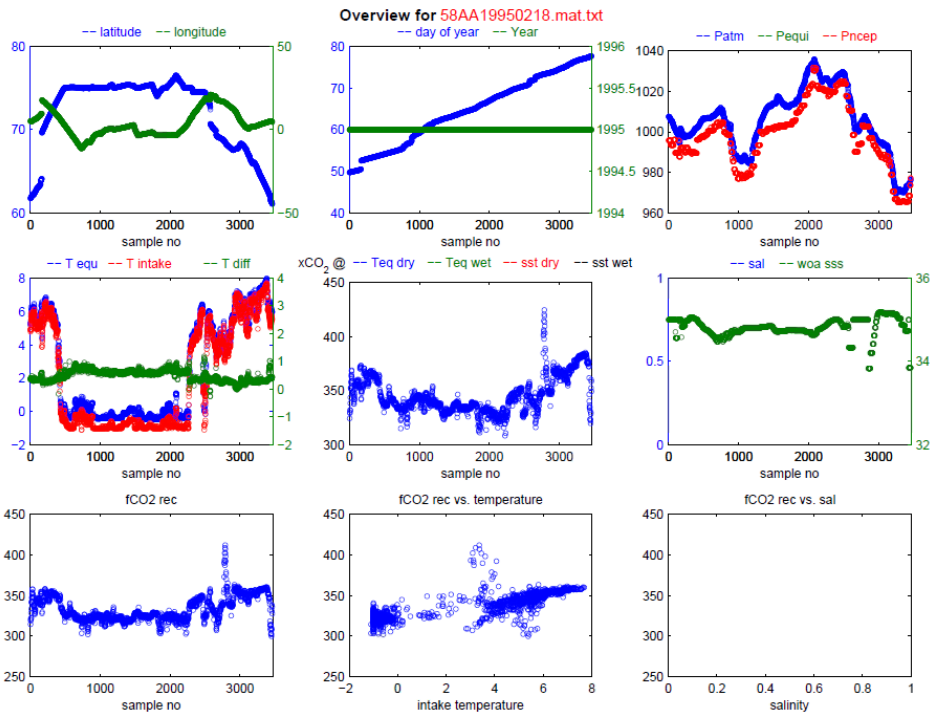


Figure 2. Example of thumbnail plots to be added to LAS QC system for data preview prior to submission

### 3.3 Crossover discussion

In the afternoon, after the workshop groups, we had a discussion regarding Crossovers. Are suggested the following:

- Figure out station pairs for each crossover
- Calculate delta of  $xCO_2$ ,  $fCO_2$ , sst and equilibrator pressure between each station pair
- Add salinity as a diagnostic parameter for cross-overs
  - Dorothee says this is especially important in coastal regions and near sea ice
  - salinity is a useful indicator of spatial variability in the data
- Print a table with time stamps and data for each pair to screen, or allow user to export such a table (to allow user to inquire the exact data involved)
- Print the weighted mean and standard deviation of the station pair differences to screen.
- For visualization, plot the differences of each pair as a function of time-distance
- Add a statement on the visualization screen as to whether the crossovers fall within the defined limits
- For crossover visualizations, it would be nice to limit plots to only crossover pairs.
  - Dorothee makes a good suggestion that instead of just showing crossover pairs, show the crossover pairs in one color (ie, blue and red) and the remaining data points in lighter colors (light blue and light red).

The LAS QC team will attempt to meet those suggestions and report back if issues prevent us from doing so. Rik and others also expressed a desire to see low quality crossovers, in addition to high quality ones. The LAS team will take this under consideration for v4.

## 4 Day 3 Activities

### 4.1 Roadmap forward



The automation team agreed that the below schedule is one we can commit to:

- V3 schedule
  - QC in progress now
  - QC Ends mid Jan
  - Release September 2015
- V4 schedule
  - Data ingestion and QC begins: Feb 2015
  - Data ingestion ends: Jan 2016
  - QC ends: April 2016
  - Release: June 2016
- V5 schedule (first annual release)
  - Data ingestion & QC ongoing
  - Data ingestion for v5 inclusion ends: 31st Jan 2017
  - QC ends March/April 2017
  - Release 30th June 2017

It is important to note that starting with version 4, data ingestion and QC can be done simultaneously.

#### 4.2 Integration of 2nd level QC into SOCAT workflow

- The creation of the 2nd level QC plots that Are and Denis showed us and that we agreed upon adding to the QC system in day 1 will be generated from the SOCAT dashboard.
- The Preview button will be relabeled as “Preview before submission” and will either generate the plots on the fly, or generate and package up and send to PI.
  - This will largely depend on how quickly the plots can be generated.
- Once the PI is satisfied with the data, they can then hit “Submit for QC” and the data will be made available for all SOCAT QC’ers

#### 4.3 Discussion SOCAT data management paper

It was suggested in Bergen, and agreed, that the automation team should put together a paper describing the SOCAT data management framework and submit to a data journal. In this final session of the meeting, we brainstorm topics that should go into the paper. These ideas are listed below.

- philosophy of why we are building the framework - requirements
  - discussion of SOCAT process prior to automation
  - Briefly discuss aim of SOCAT - collating data into easily accessible form
    - developed through SOCAT community consent
    - grass roots effort from SOCAT community to improve data quality
  - Need to streamline this process to reduce burden on data managers/QCers
  - Promote submission of high-quality data - assist PIs in this process through automatic QC
  - quick turnaround for a high quality data product
- discussion of data workflow
- discussion of technical implementation
  - ongoing ingestion and QC at the same time
- discussion of sanity checker implementation
- discussion of 2nd level QC implementation
- discussion of standardisation of metadata/documentation
- standardization of data formats
- leveraging community standards and tools for data and metadata to improve interoperability
  - for data access of QC’d data
  - use of existing tools allows for building the system at low cost

- potential for reuse of framework in other science communities
- System designed in consultation with scientists (i.e. users) at all stages - both contributors and end users
- Discuss who is using the system
- future development
  - process of annual release of high quality data
  - extensibility for additional parameters, etc

Kevin O'Brien will get a first draft of this paper out for comment mid-2015.

#### **4.4 Technical discussions on communication between the Sanity Checker and the Dashboard (Steve, Karl)**

Karl and Steve spent some time exploring the method of communication between the Sanity Checker and the Dashboard, particularly with respect to the propagation of error and warning messages. The current system is inadequate to allow the Dashboard to deal with these messages effectively. Steve will overhaul the messaging system in the Sanity Checker to better match the needs of the Dashboard.

#### **5. Meeting Adjourned**

Kevin thanks all participants for a very productive and enjoyable meeting. All participants second the sentiment, and the meeting is adjourned. There are still a few donuts remaining that are donated to the general PMEL population.