

## Summary of SOCAT Automation Planning Meeting

10 and 11 May 2012,

National Oceanographic and Atmospheric Administration - Pacific Marine Environmental Laboratory,  
Seattle, USA

*Rapporteur: Dorothee Bakker, a scientist trying to make sense of a complex automation discussion*

**Participants:** NOAA-PMEL - Steve Hankin (meeting chair, LAS), Kevin O'Brien (LAS), Heather Koyuk (LAS), Cathy Cosca (SOCAT North Pacific), Simone Alin (10 May, SOCAT coastal co-lead), Chris Sabine (10 May morning, SOCAT global); **University of Bergen, Bjerknes Centre, Norway** - Benjamin Pfeil; **Carbon Dioxide Information Analysis Centre** - Alex Kozyr; **NOAA-Atlantic Oceanographic and Meteorological Laboratory, USA** - Denis Pierrot (SOCAT global & Atlantic); **University of East Anglia, UK** - Dorothee Bakker (SOCAT global chair & Southern Ocean), Stephen Jones.

An agenda (set prior to the meeting) is in Appendix 1. We did not strictly follow the agenda, but eventually discussed most items on it.

### 1) Update SOCAT ESSD articles for SOCAT version 1.5 (Benjamin Pfeil)

Both SOCAT ESSD articles were sent around to regional group leads (plus Mario Hoppema) in mid-March. Nathalie Lefèvre and Mario Hoppema have provided comments on the SOCAT ESSD article by Pfeil et al. A revised version of both articles will be sent around to all co-authors by 18 May for prompt turn-around. Both ESSD articles will be submitted well before 1 July 2012 (IPCC deadline). Dorothee has offered her help. **Action item:** Benjamin to incorporate comments on ESSD article promptly and to send around for prompt comments by all co-authors. Dorothee to help. Submission of both ESSD articles BEFORE 1 July.

### 2) Update on SOCAT version 2 (Benjamin Pfeil)

Data submission for SOCAT version 2 was closed on 31 December 2011. Benjamin has to date incorporated 1000 additional cruises or 4 million CO<sub>2</sub> data with some data as recent as December 2011 for SOCAT version 2. The new data include many Arctic data and many data by European data PIs.

Changes in SOCAT version 2 (relative to version 1):

SOCAT version 2 will include seconds.

All expocodes will have exactly 12 characters.

Benjamin will also include extra columns (for e.g. absolute wind speed), as reported by the PI, even if these data will not be part of the SOCAT version 2 release.

For clarification:

SOCAT only includes CO<sub>2</sub> (as xCO<sub>2</sub>, pCO<sub>2</sub>, fCO<sub>2</sub>) measurements from the upper ocean (e.g. 5 m depth);

SOCAT includes CO<sub>2</sub> (as xCO<sub>2</sub>, pCO<sub>2</sub>, fCO<sub>2</sub>) data measured on ships and moorings.

SOCAT does NOT include CO<sub>2</sub> data calculated from other carbonate parameters;

SOCAT does NOT include SeaSoar data from variable depth.

Both Benjamin (Matlab) and Steve and Heather (LAS) will assess that data from version 1 will remain unaltered (even if seconds are added and Matlab scripts are run).

On the LAS earlier QC results will be visible for cruise updates.

Quality control of Arctic data version 2: Cathy Cosca and Simone Alin agree to ensure quality control of data in the Pacific sector of the Arctic Ocean for SOCAT version 2. They may contact data PIs, notably a colleague in Alaska and Helmuth Thomas (Dalhousie University).

Time table SOCAT version 2:

- Benjamin will send Heather a small number of example files asap.
- Benjamin will complete his work and send the data to PMEL by the end of May 2012.
- Steve, Kevin, and Heather will ingest the data into the LAS in June 2012.
- SOCAT participants will carry out data QC in July, August and September 2012.
- Release of SOCAT version 2 in late December 2012.

**Action item:** Benjamin to complete incorporation of version 2 data by end of May 2012.

**Action item:** Steve, Kevin, and Heather will ingest the data into the LAS in June 2012.

**Action item:** Dorothee, the global group, and regional group leads to encourage prompt data quality control by September 2012.

### Key automation items for SOCAT version 3

**Action item:** 1 March 2013 is set as the target date for completing the automation software for SOCAT (Steve Hankin, Alex Kozyr, Kevin O'Brien, Heather Koyuk, Benjamin Pfeil).

### 3) SOCAT user interface (Steve Hankin, Kevin O'Brien, Heather Koyuk)

*With much input by Stephen Jones (see Figure 1), Denis Pierrot, and discussion by all.*

SOCAT.info will host a user interface for uploading data and metadata to SOCAT.

Figure 1 shows the design of this user interface. The user interface will:

- Enable data PIs to upload metadata;
- Enable data PIs to upload data;
- Enable SOCAT to carry out file conformance tests;
- Enable data PIs to carry out initial quality control on these data;
- Enable data PIs to download their 'original' data and metadata for submission elsewhere;
- Enable data PIs to submit their data to SOCAT and:
  - a) Make their 'original' data and metadata public via CDIAC upon submission to SOCAT (instantly),
  - b) Make their 'original' data and metadata public via CDIAC upon inclusion of these data in a SOCAT release,
  - c) Take responsibility of making their 'original' data and metadata public (via a data center(s) of choice).

(At least 1 of these 3 items has to be selected as part of data submission to SOCAT).

Data submitted to SOCAT will undergo further quality control and will be subject to SOCAT reformatting and recalculation protocols. If the data are deemed of sufficient quality, they will be included and made public in the next suitable SOCAT release. Such data and metadata in SOCAT format will be archived at Pangaea.

**12-character cruise Expocodes** will link data (at SOCAT.info) with metadata (at CDIAC OME). The first two characters of a 12-character Expocode identify the country code of the vessel (<http://www.nodc.noaa.gov/General/NODC-Archive/countrylist.txt>) and are followed by the two-character

National Oceanographic Data Center (NODC) vessel code (<http://www.nodc.noaa.gov/General/NODC-Archive/platformlist.txt>). The final 8 characters denote the starting date of the measurements of the cruise (as YYYYMMDD). For instance, 06MT19920510 means that this cruise was conducted on the German (06) research vessel Meteor (MT) and that the first measurement was reported for 10 May 1992. In case of doubt, contact the CDIAC or SOCAT data managers.

#### 4) Metadata creation and management (OME, Alex Kozyr)

*Input by Are Olsen on tick boxes for key parameters (via Benjamin Pfeil);*

*Input by Cathy Cosca on project report forms.*

*Also see summary of key items by Steve Hankin (Appendix 1, sent via email on 11 May);*

*Also see input for metadata by Dorothee Bakker (Appendix 2, sent via email on 15 May);*

The SOCAT metadata will be hosted at CDIAC in OME.

The OME metadata form will communicate with the SOCAT data via the user interface by small files (MyExpocode.xml) with the Expocode as the unique cruise identifier.

Metadata should provide all information required for SOCAT quality control.

Metadata forms will be saved and past forms will be available for future use by their creators.

As part of the automation, it will be made possible to either complete metadata forms in situ or to import metadata from the headers of data files.

A facility for uploading attachments (e.g. project cruise reports, calibration reports) to the metadata forms will be added.

**Action item:** Steve Hankin to provide feedback on OME versus LAS communication (done, Appendix 2).

**Action item:** Dorothee Bakker to provide list of metadata items (done, Appendix 3).

#### 5) Data formats and file conformance tester (Steve Hankin, Kevin O'Brien, Heather Koyuk)

The self-documenting data files will have a mandatory header, an optional header and mandatory column identifiers.

The mandatory header will contain information essential for creating the Expocode for the data.

The optional header will contain metadata for automatic upload to the OME metadata form.

The column identifiers will enable identification of the column contents, as well as units. Ideally most columns can be identified automatically by standardization of the column identifiers. A user interface will further assist in identifying column contents.

Acceptable file formats will be CSV (comma) or tab separated text.

SOCAT version 3 will include CO<sub>2</sub> in marine air, as measured in on cruises.

SOCAT version 3 will not specifically include pH (total pH scale), DIC (dissolved inorganic carbon, μmol/kg), O<sub>2</sub> (oxygen, μmol/kg), TA (total alkalinity, μmol/kg), nutrients, absolute wind speed (m/s) and wind direction (degrees) at 10 m height, but these parameters will be kept prior to ingestion by the LAS, if they are reported.

Mandatory header

/\*

ID Country Vessel (2 character, <http://www.nodc.noaa.gov/General/NODC-Archive/countrylist.txt>)

Vessel ID (2 character, <http://www.nodc.noaa.gov/General/NODC-Archive/platformlist.txt>)

\*/

Optional header (to also be included in the above /\* \*/ brackets):

Start date (cruise or data set, for Expocode);

Metadata for automatic upload to the OME metadata form (with standardized identifiers).

**Action item:** Example files of acceptable data formats (notably date, time, latitude, longitude, essential contents) will be provided by Benjamin and others to Steve Hankin, Kevin and Heather.

**6) Initial Quality Control by the data PI prior to submission (Steve Hankin, Kevin O'Brien, Heather Koyuk)**

This topic was discussed briefly. The user interface will create a series of property-property plots of the data, such that the data PI can carry out initial quality control prior to submission of the data to SOCAT. One suggestion is to create property-property plots similar to output of existing programs (in Matlab) by Denis Pierrot, Ute Schuster and others.

**7) Submission of data and metadata to SOCAT (Steve Hankin, Kevin O'Brien, Heather Koyuk)**

Once the data and metadata are complete, the file conformance tests come out satisfactorily and the data PI has carried out initial quality control of the data, he/she may proceed to submission of the data and metadata. At this stage the data PI can download their 'original' data and metadata for his/her records and for submission elsewhere.

As part of the submission process to SOCAT, the PI will need to select at least 1 of these 3 options:

- a) To make the 'original' data and metadata public via CDIAC upon submission to SOCAT (instantly),
- b) To make the 'original' data and metadata public via CDIAC upon inclusion of these data in a SOCAT release,
- c) To take responsibility of making the 'original' data and metadata public (via a data center(s) of choice).

The above approach should accommodate the project commitments of individual scientists globally, while enabling these scientists to decide upon submission via CDIAC (options a & b), whenever this does not contravene their project commitments. It will ensure safe storage and archiving at CDIAC of a substantial part of the 'original' data entering the SOCAT data set (options a & b).

**8) Quality control by SOCAT participants**

This item was not discussed.

The SOCAT quality control interface will be redesigned for version 3, according to Steve Hankin and Kevin O'Brien.

**9) Versioning of SOCAT data output files and doi-numbers (Benjamin Pfeil)**

All SOCAT output files are and will be stored at Pangaea.

Each SOCAT release will be given a doi-number and a version number (e.g version 1.5).

New output files and revised output files will be given a doi-number, as they are released as part of a specific SOCAT version.

The Pangaea entry to each SOCAT data output file will have a link (a 'key') to all the SOCAT versions that this file is part of.

A comment will be added to the Pangaea entry for each SOCAT output file similar to 'Note, This file has been revised, see xxx', as a revised version of the file is released.

E.g. a (hypothetical) SOCAT output file may be part of versions 1.5, 2 and 3, while a revision to the output file might be part of versions 4 and higher.

**10) SOCAT meeting of regional group leads on versions 2 and 3 in Japan in early July 2012**

This automation meeting was partly intended as preparation of the SOCAT meeting in Japan. Several of the meeting participants will attend the Japan meeting, notably Kevin O'Brien (LAS), Cathy Cosca (SOCAT North Pacific), Benjamin Pfeil; Alex Kozyr; Denis Pierrot (SOCAT global & Atlantic); Dorothee Bakker (SOCAT global chair & Southern Ocean), Stephen Jones. Dorothee will create a draft agenda for the Japan meeting. Kevin, Alex, Benjamin, Dorothee et al. will present relevant parts of SOCAT automation plans at the Japan meeting for approval by the SOCAT community representatives.

**Action item:** Dorothee, Alex, Kevin, Benjamin et al. to prepare effective and clear presentations on automation for version 3 for approval by the participants of the Japan meeting in early July 2012.

**Action item:** Dorothee to create a draft agenda for the Japan meeting.

**Action items:**

**Action item 1:** Benjamin to incorporate comments on ESSD article promptly and to send around for prompt comments by all co-authors. Dorothee to help. Submission of both ESSD articles BEFORE 1 July.

**Action item 2:** Benjamin to complete incorporation of version 2 data by end of May 2012.

**Action item 3:** Steve, Kevin, and Heather will ingest the data into the LAS in June 2012.

**Action item 4:** Dorothee, global group, and regional group leads to encourage prompt data quality control by September 2012.

**Action item 5:** Steve Hankin to provide feedback on OME versus LAS communication (done, Appendix 2).

**Action item 6:** Dorothee to provide list of metadata items (done, Appendix 3).

**Action item 7:** Example files of acceptable data formats (notably date, time, latitude, longitude, essential contents) will be provided by Benjamin and others to Steve Hankin, Kevin and Heather.

**Action item 8:** Dorothee, Alex, Kevin, Benjamin et al. to prepare effective and clear presentations on automation for version 3 for approval by the participants of the Japan meeting in early July 2012.

**Action item 9:** Dorothee to create a draft agenda for the Japan meeting.

**Action item 10:** To prepare for the V2 QC process, Heather will make the following additions to the LAS:

1. username/password authentication, so LAS "knows" unambiguously who is entering the QC
2. a checkbox to control whether the user sees V2 QC flags, only, or V2 + V1.5
3. a checkbox that allows the user to get the full extent of the data from cruises, even when the subset of cruises has been chosen based upon passing through a particular region

**Action item 11:** Steve Jones will be developing the "sanity checker" (Java code) for the automation. He will develop a comprehensive list of all SOCAT-recognized column headings for data files (parameter and units)

**Action item 12:** PMEL will develop the semi-automated Web interface to allow users to clarify ambiguities in the column headings of their files. PMEL will provide to S.J. the specification that communicates the column headings to the sanity checker.

**Action item 13:** PMEL will send a list of technical topics to Alex re: what more will be needed to harmonize data and metadata entry in the automated system that is distributed between ORNL and SOCAT.info. Together we will consider whether ORNL's programmer should travel to PMEL for a small technical meeting to find solutions to these topics.

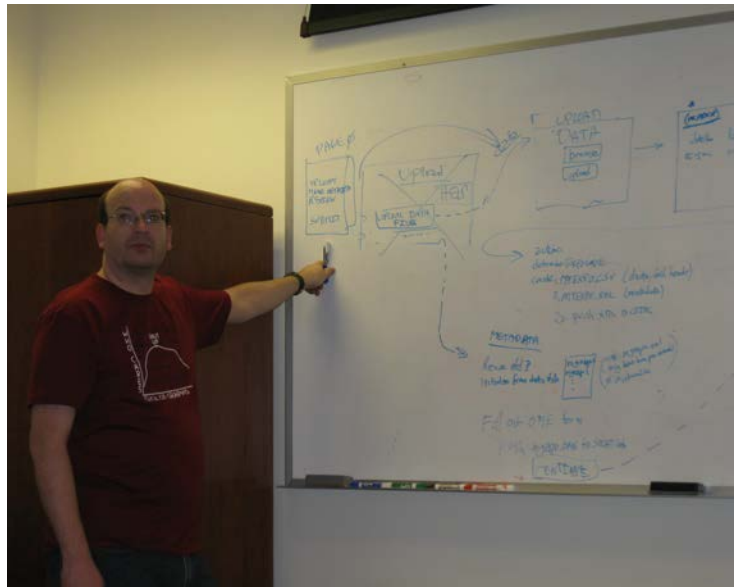
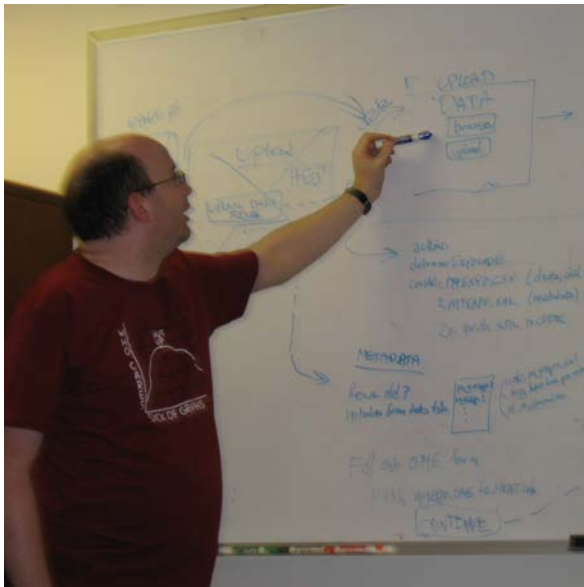
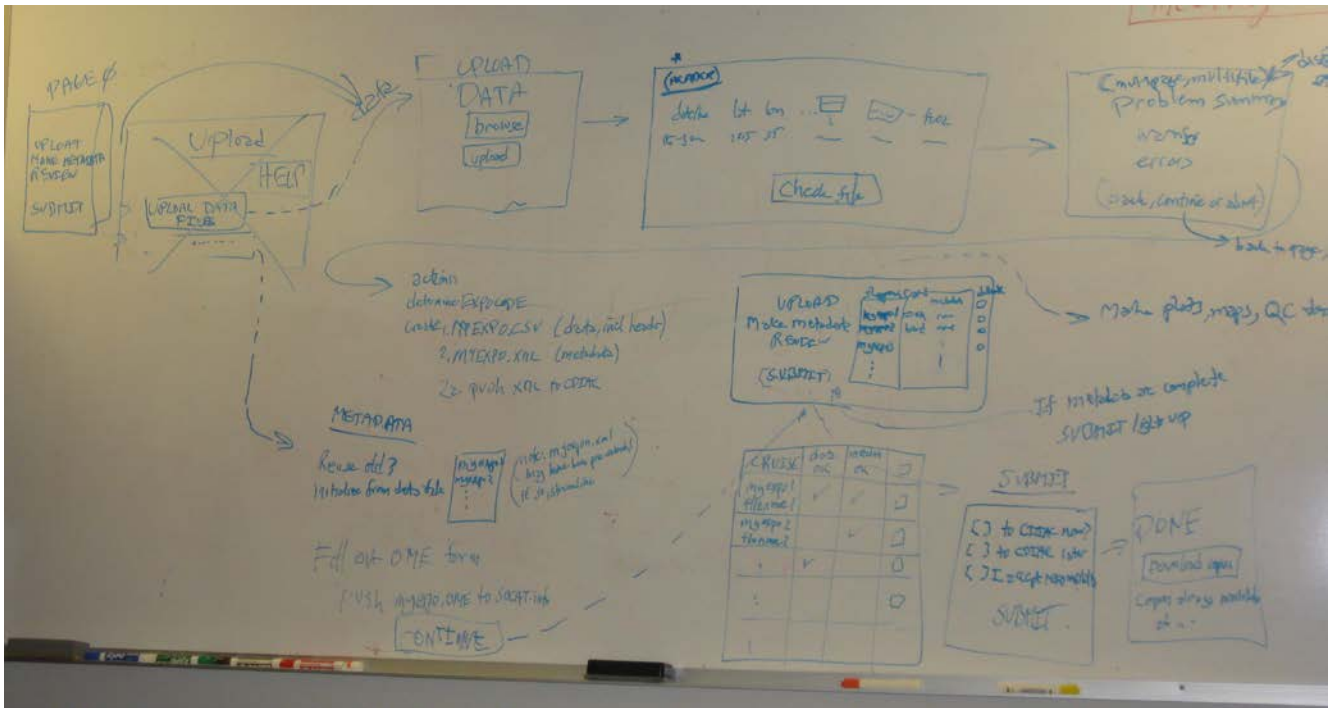


Figure 1. The SOCAT automation user interface by Steve Hankin, Kevin O'Brien, Stephen Jones and Denis Pierrot (lunch time breakout group on 11 May).

**Appendix 1: Agenda** (prior to meeting):

10 May

- Welcome and logistics
- Metadata creation and management (Alex, Benjamin)
  - Critique of metadata issues / concerns from the pat (Cathy)
  - OME process for submission/updating (Alex)
  - Machine access to metadata records from the SOCAT system
  - Creation of ISO metadata for data discovery portals
  - OME metadata vis a vis institution specific metadata?
- Creation of doi-s (Benjamin)
  - Who creates them? When? Issues of credit for doi-s?
- Format of data files (Steve Hankin, Kevin, Benjamin)
  - Standardizing variable file names
  - Metadata in file header

11 May

- Mop up left over topics from day 1
- Software design and capabilities (Steve Hankin, Heather, Kevin)
  - Agree on system requirements (see power point handout)
    - Authentication
    - File uploads
      - Expocode in header
      - File conformance tester
    - Retaining past versions during QC process
    - QC entry (suspending & removing cruises)
    - Submission of data and metadata files to the national centers
  - Review SOCAT system from version 1.5 and critique needs for improvement
- Summarize action items



**Appendix 2. Notes from Steve Hankin on metadata in OME** after the automation meeting  
(sent to Alex Kozyr by email on 11 May at ~18:00 Seattle time):

- The form entry process should ask the person entering the information for *\*their\** contact info, so you are able to contact them when problems are detected
- Expocodes should nowhere be explicitly entered by the user. Instead the user enters the i) country; ii) vessel ID; and iii) start time to be used in expocode. The OME system should create the expocode at some point after these fields have been entered and should inform the user what the code is.
- When entering the variables that require metadata, a certain list of core SOCAT variables should be pre-named in the form. Some of these variables may require special questions on the form. (Science PIs to provide you with a list of which variables these are.)
- Metadata form should provide explanations of what each field means. This documentation might be seen by hovering the mouse over the name of the field in the form ... or by clicking on a "?" button; or by some other method
- When starting up a new metadata session the user should be able to choose whether to "Edit previously entered metadata" or "Start a new metadata entry by copying a previous metadata". If a "copy" is chosen, then the start date of the cruise should be removed, which will cause the expocode to be unknown.
- Need to be able to add "attachments" to a metadata entry. Preferably the user should be able to enter a "description" field to go with each attachment.
- For SOCAT a delayed approval process on a new metadata record would be helpful. In other words, the system would accept the new metadata record right away, but flag somewhere in the metadata that the record is provisional. The delayed approval would remove the stamp of being provisional.
- The socat.info system will need to be able to "push" XML files to ORNL that contain information that will later be used to initialize the metadata form with information already known about particular cruises. Details of this need to be worked out:
  - how to push this document
  - the format/schema of the information in the document
  - how to trigger the OME system to initialize from a particular one of these documents

- Steve

### Appendix 3. Input for OME Underway metadata form by Dorothee Bakker (as sent by email to Alex Kozyr on 15 May 2012)

#### Metadata Creator Information

email xxx

#### Investigator

name xxx

Organisation xxx

Address xxx

Phone xxx

Email xxx

#### Dataset Info

Funding information (funder, project name, grant number) xxx

#### Submission dates

Initial submission yyyy/mm/dd

Revised submission yyyy/mm/dd

#### Cruise information for the Expocode

ID Country Vessel (<http://www.nodc.noaa.gov/General/NODC-Archive/countrylist.txt>) xx ONLY

Vessel ID (<http://www.nodc.noaa.gov/General/NODC-Archive/platformlist.txt>) xx ONLY

Start date (cruise or data set, as in Expocode) yyyyymmdd ONLY

Cruise Expocode (as Expocode in data file, e.g. 06MT19920510) xxxxyyyymmdd ONLY

#### Other Cruise info

Survey type (e.g. VOS lines, Research cruise, moored buoy, drifting buoy) xxx

Vessel name xxxxx

Cruise Info (e.g. SAVE, TTO-NAS, SOIREE, AMT08, Antares) xx

Cruise name (including Leg) (E.g. ANTV-2, Biscay\_979815C, D198) xxx

End date (of cruise or data set) yyyyymmdd

Ports of call (one per line) xxx

#### Geographical coverage

Geographical region text box

#### Bounds:

Westernmost Longitude (to nearest degree west of area, e.g. 10W) xxx

Easternmost Longitude (to nearest degree east of area, e.g. 2W). xxxxx

Northernmost Latitude (to nearest degree north of area, e.g. 60S). xxx

Southernmost Latitude (to nearest degree south of area, e.g. 65S) xxxxx

#### Variables Info

Type of reported CO2 data (xCO2, pCO2, fCO2) tick boxes (below)

xCO2water\_equ\_dry tick box with unit  $\mu\text{mol/mol}$

xCO2water\_SST\_dry tick box with unit  $\mu\text{mol/mol}$

pCO2water\_equ\_wet tick box with unit  $\mu\text{atm}$

pCO2water\_SST\_wet tick box with unit  $\mu\text{atm}$

fCO2water\_equ\_wet tick box with unit  $\mu\text{atm}$

fCO2water\_SST\_wet tick box with unit  $\mu\text{atm}$

xCO2air\_dry tick box with unit  $\mu\text{mol/mol}$

pCO2air\_wet tick box with unit  $\mu\text{atm}$

fCO2air\_wet tick box with unit  $\mu\text{atm}$

xCO2air\_dry\_interpolated tick box with unit  $\mu\text{mol/mol}$

pCO2air\_wet\_interpolated tick box with unit  $\mu\text{atm}$

fCO2air\_wet\_interpolated tick box with unit  $\mu\text{atm}$

Unit of reported CO2 data xxx

Total reported CO2 variables in the data set xxx

#### Surface water CO2 method

##### Sampling & Equilibrator design

Depth of seawater intake (m) xxx

Location of seawater intake text box

Equilibrator type text box

equilibrator volume (L) xxx

Water_flow_rate (L/min)	xxx
Headspace_Gas_flow_rate (L/Min)	xxx
Equilibrator Vented (Yes/no)	xxx
Drying method for CO2 in water and extent of drying	text box
Additional comments on equilibration	text box
<b>CO2 in marine air method</b>	
Measurement of CO2 in marine air (yes/no) & frequency	text box
Location and height of marine air intake	text box
Drying method for CO2 in air and extent of drying	text box
<b>CO2 sensors</b>	
Measurement method CO2	xxx
Manufacturer of CO2 sensor	xxx
Model of CO2 sensor	xxx
Frequency of CO2 measurements	xxx
Accuracy of CO2water (specify parameter and unit)	xxx
Precision of CO2water (specify parameter and unit)	xxx
Accuracy of CO2air (specify parameter and unit)	xxx
Precision of CO2air (specify parameter and unit)	xxx
CO2 calibration gases: manufacturer, mixing ratio of CO2 standards, calibration	text box
Calibration of CO2 calibration gases (traceability to an international scale)	text box
Additional comments on CO2 analysis	text box
CO2 method references (publications describing method)	text box
<b>Sea Surface Temperature (SST)</b>	
Location and depth of SST sensor	text box
Manufacturer SST sensor	xxxx
Model SST sensor	xxxx
Accuracy SST (degrees Celsius)	xxxx
Precision SST (degrees Celsius)	xxxx
Calibration of SST (traceability to an international scale)	text box
Additional comments on SST analysis	
<b>Equilibrator Temperature (Tequ)</b>	
Location of Tequ sensor	text box
Manufacturer Tequ sensor	xxxx
Model Tequ sensor	xxxx
Accuracy Tequ (degrees Celsius)	xxxx
Precision Tequ (degrees Celsius)	xxxx
Calibration of Tequ (traceability to an international scale)	text box
Average warming of water with standard deviation (degrees Celsius)	
Additional comments on Tequ analysis	
<b>Equilibrator Pressure (Pequ)</b>	
Location of Pequ sensor	text box
Manufacturer Pequ	xxxx
Model Pequ	xxxx
Accuracy Pequ (specify unit)	xxxx
Precision Pequ (specify unit)	xxxx
Calibration of Pequ (traceability to an international scale)	text box
Additional comments on Pequ analysis	
<b>Atmospheric Pressure (sea level) (Patm)</b>	
Location and height of Patm sensor	text box
Manufacturer Patm	xxxx
Model Patm	xxxx
Accuracy Patm (specify unit)	xxxx
Precision Patm (specify unit)	xxxx
Calibration of Patm (traceability to an international scale)	text box
Additional comments on Patm analysis	
<b>Sea Surface Salinity (SSS)</b>	
Location of SSS sensor	text box
Manufacturer of SSS sensor	xxxx
Model of SSS sensor	xxxx
Accuracy of SSS (no unit)	xxxx

Precision of SSS (no unit)	xxxx
Calibration of SSS (traceability to an international scale)	text box
Additional comments on SSS analysis	
<b>Other sensors (optional)</b>	
Total of additional sensors in the data set	xxx
Manufacturer of additional sensor 1 to x	xxx
Model of additional sensor 1 to x	xxx
Accuracy of additional parameter 1 to x (specify unit)	xxx
Precision of additional parameter 1 to x (specify unit)	xxx
Calibration of additional sensor 1 to x (traceability to an international scale)	text box
Additional comments on other sensor 1 to x	text box
Additional Information	text box
Data Set References (Publications describing data set)	text box
Citation (How to cite this data set)	text box
Append measurement or calibration reports here (optional).	