

How to submit a dataset to SOCAT ?

Get a quick overview of the procedure by watching the [videos on the website](#) (submission and metadata).

Step 1 – Pre-submission checks	2
Step 2 – Uploading a dataset	4
Step 3 – Preview data plots.....	9
Step 4 – Metadata	10
Option 1 – Metadata Editor	10
Option 2 – Own file	12
Step 5– Submitting to SOCAT and choose of the repository.....	13
Step 6 – What happens after the submission ?	14

Note : Any new datasets you submit remain yours.

Note : Data submissions are accepted throughout the year. Submitting data continuously helps reduce pressure on the submission and processing interface. However, only datasets submitted before the January deadline will be quality-controlled (QCed) and released in the same calendar year. Datasets submitted after the deadline will be processed and released in the year after.

Last updated : June, 2026

Step 1 – Pre-submission checks

The data file you upload to the Dashboard from the “Upload Data” page, must be either a file of tab-separated values (“TSV” format) or comma-separated values (“CSV” format). This data file has the general format:

- ✓ a line of metadata for each metadata item,
- ✓ a line of data column headers,
- ✓ a line of data column units (optional),
- ✓ a line of data values for each data sample

The expocode, vessel (ship) name, and investigator name(s) for the dataset must be given in either the metadata lines at the beginning or in columns with appropriate names.

The following is the start of an acceptable CSV file with units as part of the column name header line. The data header and data values lines were truncated to make it easier to see the format.

```
Expocode: 33AT20120417
Ship: Atlantis
PI: Wanninkhof, R.
JD_GMT, DATE.UTC__ddmmyyyy, TIME.UTC_hh:mm:ss, LAT_dec_degree, LONG_dec_degree, ...
110.79219, 19042012, 19:00:45, 12.638, -59.239, ...
110.79391, 19042012, 19:03:14, 12.633, -59.233, ...
110.79564, 19042012, 19:05:43, 12.628, -59.228, ...
110.79736, 19042012, 19:08:12, 12.622, -59.222, ...
110.79910, 19042012, 19:10:42, 12.617, -59.216, ...
```

The following is the start of another acceptable CSV file which has units as a second header line and uses a slightly different format for the lines of metadata.

```
Expocode = 33AT20120417
Vessel Name = Atlantis
Investigator = Wanninkhof, R.
JD_GMT, DATE.UTC, TIME.UTC, LAT, LONG, ...
, Jan1=1, ddmmyyyy, hh:mm:ss, dec.deg., dec.deg., ...
110.79219, 19042012, 19:00:45, 12.638, -59.239, ...
110.79391, 19042012, 19:03:14, 12.633, -59.233, ...
110.79564, 19042012, 19:05:43, 12.628, -59.228, ...
110.79736, 19042012, 19:08:12, 12.622, -59.222, ...
110.79910, 19042012, 19:10:42, 12.617, -59.216, ...
```

Object	Accepted terms
Expocode	✓ expocode ✓ cruise expocode
Ship	✓ ship ✓ ship name ✓ vessel ✓ vessel name
Investigator name	✓ investigator ✓ investigators ✓ investigator name ✓ investigator names ✓ PI ✓ PIs ✓ PI name ✓ PI names

Tags for metadata items are case insensitive. The tag is followed by either a colon or equals sign, which can have spaces around them. For datasets with multiple investigators, put all names on one metadata line and separate the names with semicolons. If metadata is contained in data columns, the header for the data column should be one of these tags (left table).

The first line of comma-separated (for CSV format) or tab-separated (for TSV format) values without any blank or numeric entries will be considered the line of data column headers¹. If the line following the header line has no numeric entries, it will be treated as a line of data column units; otherwise, it will be the first data sample line. Any lines with no values (all blank values) will be ignored.

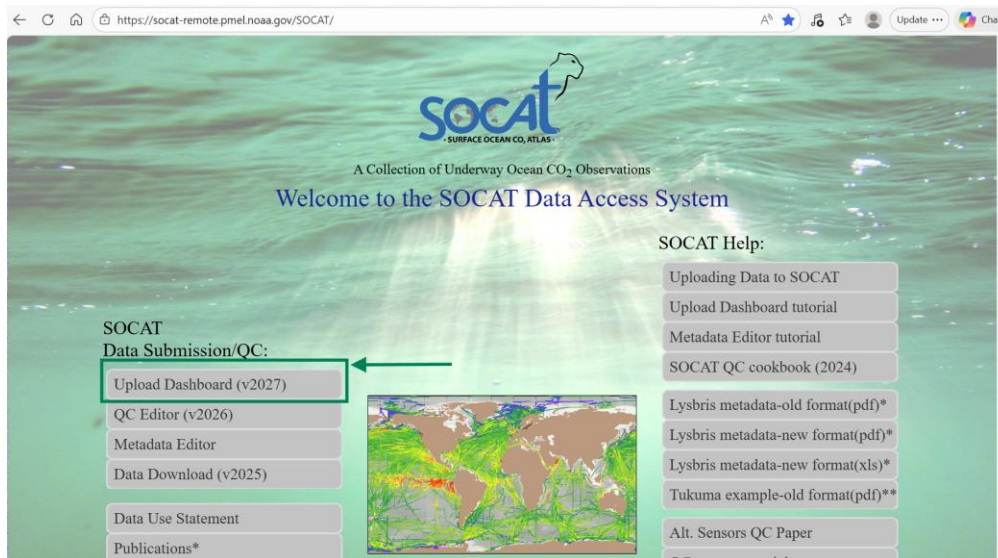
When you upload your data file, the Dashboard will examine it and attempt to identify the column header names and units (see [Step 1 – Pre-submission checks](#)).

¹ Because of this, metadata lines can have blank entries (such as ‘,,,,’ in CSV format files) following the metadata value.

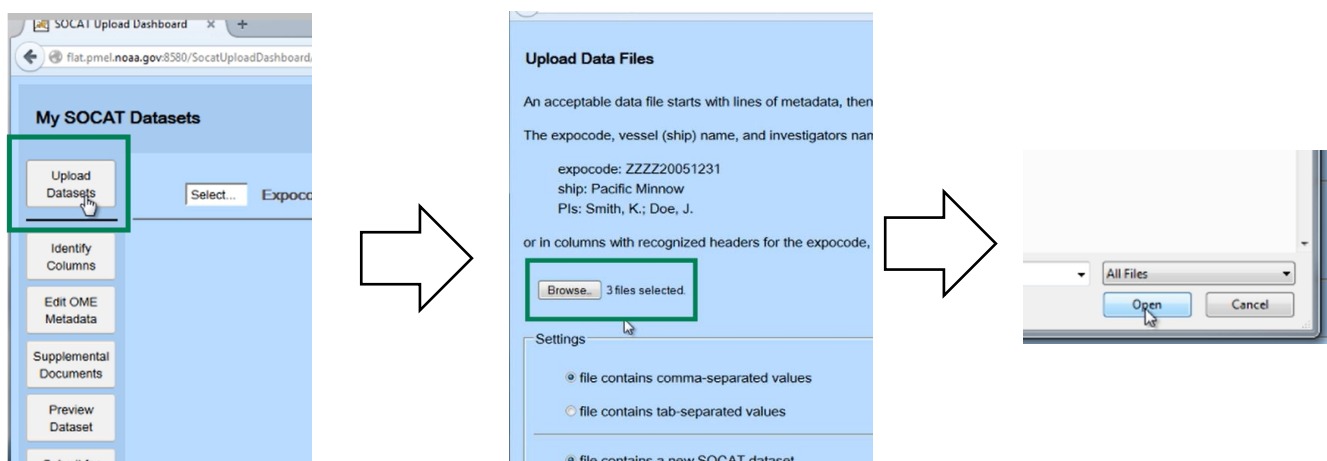
Step 2 – Uploading a dataset

First, you need an account to log in to the SOCAT Upload Dashboard. If you don't have one, please request an account to oar.pmel.socat.support@noaa.gov, and wait for confirmation.

With your account, please log in to access the SOCAT Data Access System (LAS) : <http://socat-remote.pmel.noaa.gov/>. On the left, click on “**Upload Dashboard** (v....)”.



In the menu on the left, select “**Upload Dataset**”. On the new page, click “**Browse**” and navigate to the folder containing your dataset(s) (accepted formats: csv or tsv). After selected the relevant document(s), click “**open**” in the lower-right corner of the window.



The main page of the Dashboard lists all cruises that belong to you². If you have not uploaded any datasets, your list may be empty, or it may contain SOCAT v3 (and earlier) datasets that were marked as belonging to you.

The dataset belongs to someone else ?

If there are mistakes, please notify socat.support@noaa.gov with information about the mistaken ownership. You can also reassign ownership of any datasets in your list using the “Change Datasets Owner” button. This can be used to let a colleague view your private data, or correct mistaken ownership. Datasets from previous versions of SOCAT can be hidden from your displayed list (using the “Hide Datasets” button) to reduce the number of datasets shown. If later you need to work with the dataset, for instance to fix mistakes in the data or metadata, you can use the “Show Datasets” button to return it to your list.

How to find the expocode of you dataset ?

All datasets are managed using their expocode, which is a unique³ identifier of the dataset. It is important you know the expocodes of your datasets, and that expocodes of new datasets are correct.

The 12-character expocode is the NODC code⁴ for the vessel carrying the instrumentation followed by the numeric year, month, and day of departure. For example, 49P120101218 indicates a cruise on the Japanese (49) ship of opportunity Pyxis (P1) with the first day of the cruise on 18 December 2010.

Tip : You can select/upload one or several cruises at the same time.

On the dashboard, you can now preview the data. Click “**Advanced options**”, scroll down and select “**Preview Data File**”. The first lines of your dataset should be displayed. If the preview is correct, please proceed with the next steps. ***If it is incorrect, please review your dataset locally on your computer***, apply the necessary corrections, and upload the updated file to the Dashboard. Repeat this process until the preview displays correctly.

² If you are registered as a “manager” in the Dashboard, you can also view and edit datasets of users whom you manage or other managers you collaborate with.

³ Sometimes two datasets, particularly for moorings which share the same NODC code, will have the same 12-character expocode. In these cases, SOCAT currently adds a hyphen and a number to give unique 14-character expocodes. Better solutions to this problem are being discussed.

⁴ See <http://www.nodc.noaa.gov/General/NODC-Archive/platformlist.txt>

Upload Data Files

An acceptable data file starts with lines of metadata, then has a line of column headers, and then the data rows.

The expocode, vessel (ship) name, and investigator names must be provided in the first three columns. For example:

```
expocode: ZZZZ20051231
ship: Pacific Minnow
PIs: Smith, K.; Doe, J.
```

or in columns with recognized headers for the expocode, ship name, and investigator names.

Browse... 3 files selected.

Settings

- ☒ file contains comma-separated values
- ☐ file contains tab-separated values
- ☒ file contains a new SOCAT dataset
- ☐ file replaces an existing SOCAT dataset

[Advanced Options](#)

Upload Cancel

☒ file contains a new SOCAT dataset

☐ file replaces an existing SOCAT dataset

Advanced Options

Select a character set encoding for this file.

- If you are unsure of the encoding, UTF-8 is the safest choice.
- The main differences in UTF-8 and ISO-8859-1 are that UTF-8 can handle a much larger range of characters, and it is the standard encoding for the web.
- Use UTF-16 only if you know your file is in that encoding.
- Use the Windows encoding only for files produced by Windows applications.
- The preview button will show the beginning of the file.

File encoding: UTF-8

Preview Data File

If the preview appears correct, scroll back up and select the appropriate options describing the dataset format (csv or tsv) and indicate whether this dataset already exists in the database or represents a new submission. Click “upload”.

Browse... 3 files selected.

Settings

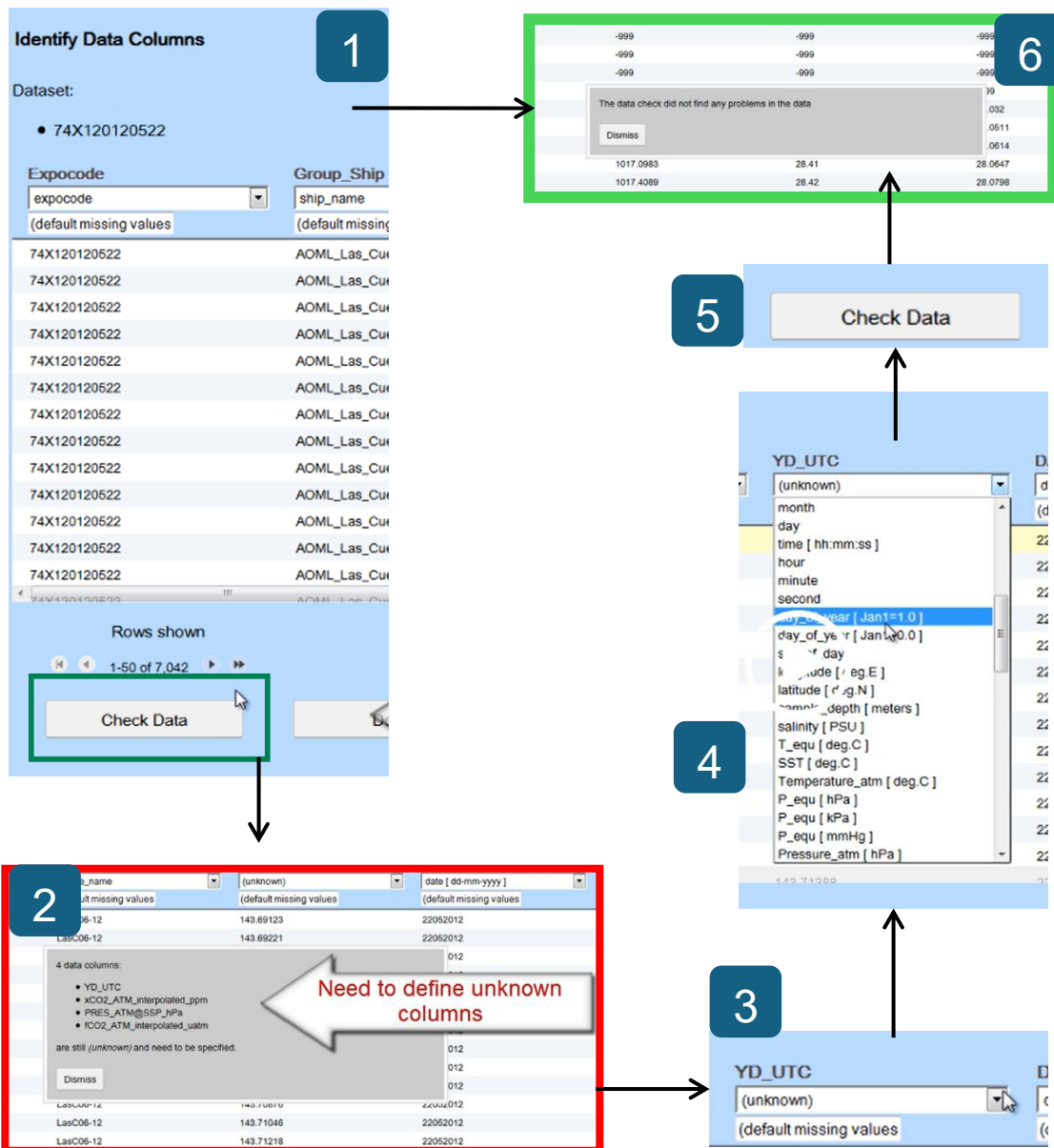
- ☒ file contains comma-separated values
- ☐ file contains tab-separated values
- ☒ file contains a new SOCAT dataset
- ☐ file replaces an existing SOCAT dataset

[Advanced Options](#)

Upload Cancel

You will be redirected to a page called “Identify Data Columns”. The objective is to provide known names for the columns, so that the SOCAT system can harmonize them across all datasets. The data columns in the dataset(s) you upload will be identified from the data column headers, and the data will be checked for obvious errors.

By clicking “**Check Data**” in the lower left corner of the screen (1), the system will check the name of each column and automatically identify those that are unknown to it. Click “**Dismiss**” (2) and proceed as follows: for each unknown column, click on the arrow on the right (3) and select the appropriate pre-registered name (4). Repeat this step until all columns have known names (3 and 4). Click “Check Data” again (5). If all columns are now known, the system will display the message “Data check did not find problems in the data”. Click on “Done”.



Missing name possibilities?

Non-standard missing values for any data columns can also be specified. Columns of data that are not used in SOCAT at this time, such as nutrient information, can be marked as the “Other” data type. Any data column marked as the “Other” data type will be ignored for SOCAT purposes; but will remain in the original data file that SOCAT saves.

If you face undefined errors...

Note that all data samples must have a valid longitude, latitude, date, and time of measurement⁵. Furthermore, the data samples must be ordered in ascending time order - from the beginning of the cruise to the end of the cruise - with no duplication of time values. Datasets with longitude, latitude, or time errors are marked as such in your list of datasets and little can be done with these datasets other than see what problems the Dashboard discovered in the data. You should correct the errors discovered and upload the updated data to the Dashboard to continue working with such a dataset.

Smart dashboard – customized column headers

The Dashboard has an extensive listing of common column header names/units and their column types. When you identify columns that were not recognized, which were incorrectly identified, or which uses a non-standard missing value, the Dashboard saves this information as your personal customization of the Dashboard. Future uploads of data files will then use your personalized list of column headers and their types, and assign the types you last gave for these column headers. So if you use column names that the Dashboard does not recognize, you only have to identify the columns once, and the Dashboard “learns” these column names.

To assist in adding many datasets to SOCAT, you can select multiple data files when you upload data in the Dashboard. When you do this, the first data file will be presented, and you 5Because of this, metadata lines can have blank entries (such as ‘,,,,’ in CSV format files) following the metadata value. 6Common mistakes in the date format include the ordering of the day, month, and year as well as whether the year is two digits or four. For full-date formats without separators between the day, month, and year, the exact number of digits for that format must be present. Any standard separator, not just the one shown, is acceptable in date formats with separators. need to identify any unknown data columns and check the data, as is done with a single data file upload. However, when you are done examining and checking that first data set, the Dashboard will then automatically identify the data columns and check the data in the remaining datasets without any further interaction from you. If you are consistent in the naming of your data columns, the dashboard “learns” any custom column names from your first data file and uses any of your customizations for the remaining data files. This allows you to quickly process many datasets with a minimal amount of interaction. When the Dashboard returns to your list of datasets after a multiple-file upload, check that there are few, if any, errors or warnings for each of the datasets you uploaded.

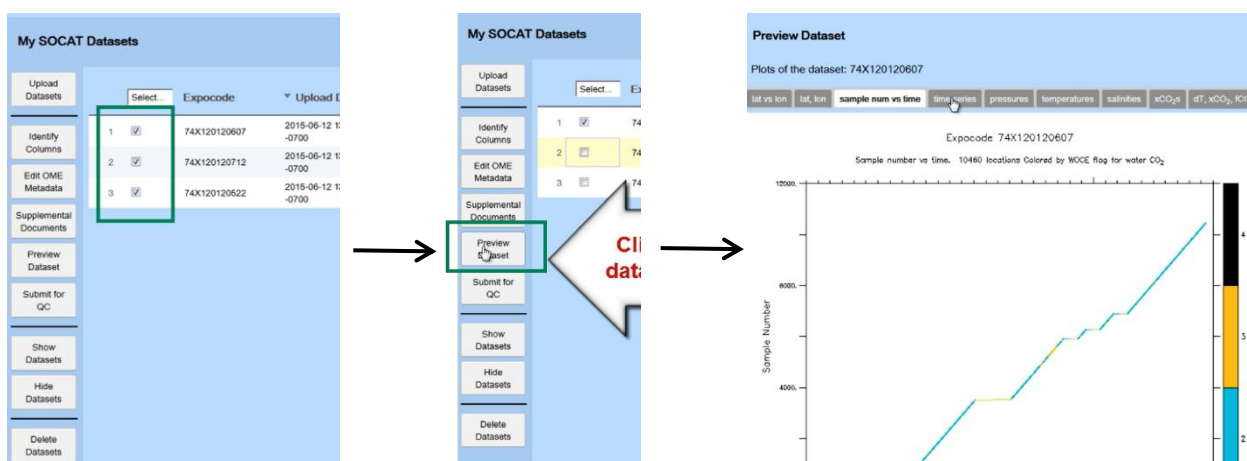
⁵ Common mistakes in the date format include the ordering of the day, month, and year as well as whether the year is two digits or four. For full-date formats without separators between the day, month, and year, the exact number of digits for that format must be present. Any standard separator, not just the one shown, is acceptable in date formats with separators.

Step 3 – Preview data plots

A set of plots of the data (the “Preview Plots”) is available to help detect inconsistencies or other errors in your data. After uploading the data for a dataset, you should examine plots of your data on the “Preview Data” page. Back on the main page of the Dashboard, please select one dataset and preview it by navigating through different pre-defined plots. In the left menu, select “**Preview Dataset**”. After a short loading period, all available pre-defined plots will appear. We invite you to navigate through them and evaluate whether aberrant values or behaviors are present in the dataset.

If you discover errors that need to be corrected, you should correct your data file and re-upload the updated dataset. Sometimes the error may just be a misidentification of a data column, its units, or its missing value. In this case, you can correct the identity of the data columns in the Dashboard without having to upload the data (see [Step 1 – Pre-submission checks](#)).

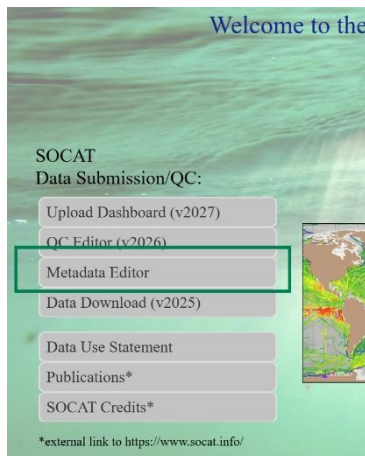
If you are satisfied with the data, please click “**Done**” at the bottom of the page.



Step 4 – Metadata

To submit a dataset to SOCAT, you also need to upload one or more metadata files describing the data and how they were obtained. This metadata should identify each of the data columns in your data file (including data columns that are marked as “Other” data types) and describe details of how the data was collected.

Option 1 – Metadata Editor



This first option is to provide metadata using the [Online Metadata Editor](#) (“OME”) for SOCAT provided by PMEL. On the main page of the Dashboard, please select “Metadata Editor” on the left.

If you use the editor for the first time, or if you use new equipment requiring different metadata, click on each category on the left and complete the form in full. If you fill in all spaces indicated in red (mandatory to submit for SOCAT), it will be sufficient metadata for the community. The first step is to enter an expocode in the matching field.

Enter the Information About this Citat

(*) Denotes a required field.

EXPOCODE *	Cruise ID
EXPOCODE	Cruise ID
Brief Project Title	

If you have already created a metadata file before (formats: XML, CSV, Excel), and you need a document quite similar (only a few changes), we invite you to click “**Choose file**” → browse on your computer, select the file → click “**open**” → “**upload**” (on the left). You will see that some information has automatically been filled in the editor.

Once you believe that all information has been filled in, press “**Validate**”. If there are missing values, they will be highlighted in red on the menu.

Once all information is there, press “**validate**”, and the system will indicate that your document is available for download. Press “**Download**”, then save it as XML file (“**Save**”). Save it on your computer.

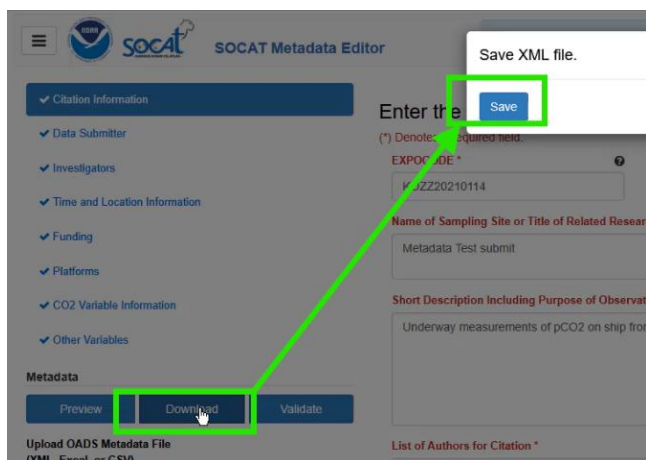
Metadata

Preview Download Validate

Upload OADS Metadata File
(XML, Excel, or CSV)

Choose file:

Upload Clear All



Some fields are empty ?

It's normal. Not all fields are required to submit data for SOCAT. However, you are always welcome to fill them up when you believe it is relevant or useful, or simply when the information is available.

This file is then uploaded to the Dashboard using the “OME Metadata” page (see below). When OME metadata are provided, the date and time of the upload are displayed in the “OME Metadata” column in the Dashboard dataset listing. An easy-to-read document will be generated from the OME XML file you uploaded, and this document will be available

for the scientists performing quality control for SOCAT. The OME XML file will also greatly simplify archiving your dataset at NCEI⁶ if you choose to do so.

My SOCAT Datasets



Upload Datasets

Identify Columns

Edit OME Metadata

Select...	Expocode	Upload Date	Data Status	OME Metadata	Supplemental Documents	QC Status	Archival
1 ☒	74X120120607	2015-06-12 13:59 -0700	No warnings	(no metadata)	(no documents)	Private	Not specified
2 ☐	74X120120712	2015-06-12 13:59 -0700	No warnings	(no metadata)	(no documents)	Private	Not specified
3 ☐	74X120120522	2015-06-12 13:59 -0700	No warnings	(no metadata)	(no documents)	Private	Not specified

Did you know you can preview the metadata file ?

By clicking “Preview” in the Metadata Editor, you can preview of the document will appear. This feature allows you to review the information provided and check for potential errors before submission.

Option 2 – Own file

For the dataset you are working on, select “**No document**” in the “Supplemental Documents” column.

My SOCAT Datasets



Upload Datasets

	Select...	Expocode	Upload Date	Data Status	OME Metadata	Supplemental Documents	QC Status	Archival	
Identify	1	☒	74X120120607	2015-06-12 13:59 -0700	No warnings	(no metadata)	(no documents)	Private	Not specified

Click on “**Browse**” and select your document (format: PDF or XML ?). Click on “**Open**” in the lower right corner of the loading window. Click on “**Upload**” and then “**Done**”.

The files you upload will be uploaded and associated with the dataset(s) you selected; the contents are not examined or modified in any way. Because of this, the uploaded files should be in standard formats that will be easy to read by scientists performing quality control for your cruise. If you wish to have your dataset archived at CDIAC, it is strongly recommended you use the OME option for metadata. The names of uploaded files, as well as the date and time of their upload, will appear in the “Supplemental Documents” column of your list of datasets in the Dashboard.

The Supplemental Documents page can also be used to update (by just uploading or deleting any supplemental documentation).

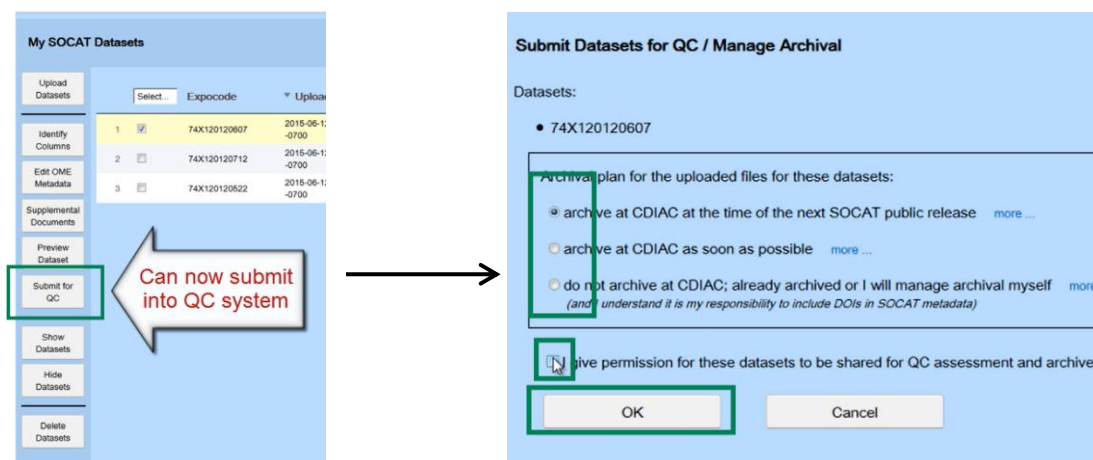
⁶ Formerly CDIAC

Step 5– Submitting to SOCAT and choose of the repository

Once Metadata has been uploaded, it's time to submit the dataset and metadata to SOCAT for the QCing step.

In the left menu of the Dashboard, select “**Submit for QC**”. You will be asked to choose between repository options. Select the one that is the most appropriate for your situation.

Please check the box allowing SOCAT to review datasets. Click “**OK**”. Your dataset is submitted.



Which option choosing ?

Since all the data used for SOCAT must be archived, you are presented with three options for archiving the dataset you uploaded to SOCAT.

1. The first and the default is to have the data archived at CDIAC when the next version of SOCAT is released. This option allows SOCAT scientists to examine your data, possibly resulting in corrections or additions to your data or metadata files, before the dataset is archived.
2. The second option for archival is to archive at CDIAC immediately. This option makes your uploaded data and metadata files available to the public as quickly as possible but will also require resubmitting these files after corrections and other updates.
3. The third archival option indicates that the dataset is already archived (such as when the dataset is already published), or that you will archive it at some standard data archival center (if you wish to archive it somewhere other than CDIAC) before the next release of SOCAT. For this option, we request that you provide the DOI or other archival reference to the dataset in the metadata you provide to SOCAT so we may properly cite your dataset within SOCAT.

Step 6 – What happens after the submission?

Once a dataset is submitted for QC, you are limited in what you can do with the dataset on the Dashboard. You can always modify your data archival option given on the “Submit for QC page”, reexamine the data-check results from the “Identify Columns” page, and re-examine the preview plots of your data from the “Preview Dataset” page. However, you will not be able to update the data or the data types while a dataset is submitted for QC.

If you need to update the data, the dataset needs to be suspended using the SOCAT QC Editor. The QC status in the Dashboard will then change to “Suspended” and you can modify the dataset as desired⁷. When you are done with your changes, you then submit your dataset for QC once again.

If you only need to update the metadata for your dataset, either to correct errors or to add missing information to improve the QC rating, you do not need to suspend the cruise. Instead, use the “Edit OME Metadata” or “Supplemental Documents” page to modify, delete, or add the metadata documents as desired. When the metadata documents are modified, the dataset will be marked as updated in the QC editor (and thus needing to be QC’ed again), but since the data has not been modified, any WOCE flags that had been assigned to the data itself will remain in place⁸.

⁷ This change in the Dashboard QC status may take up to 15 min. to occur. An improvement to have that change occur much faster is underway.

⁸ The comment associated with the update QC flag in the SOCAT QC Editor will indicate that only metadata has changed; data and WOCE flags are unchanged. This can simplify the re-QC of the dataset, but you should also add a QC comment to your dataset in the SOCAT QC Editor indicating how the metadata changed to further simplify the re-QC of the dataset.