

The SOCAT Upload Dashboard (the “Dashboard”) is used to upload new and updated datasets for submission to SOCAT. 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. A set of plots of the data (the “Preview Plots”) are available to help detect inconsistencies or other errors in your data. To submit a dataset to SOCAT, you also need to upload one or more metadata files describing the data and how the data was obtained. Future versions of the Dashboard will allow you to create and modify files of complete metadata (the “OME Metadata”) in the Dashboard, eliminating the need to create metadata files elsewhere. When you submit a dataset for QC to SOCAT, you can also request that the data and metadata you have uploaded will be archived at CDIAC.

To begin using the Dashboard, you need to login using your Dashboard username. If you do not have a Dashboard login, contact [socat.support@noaa.gov](mailto:socat.support@noaa.gov) to obtain an account. You will only be able to view and modify datasets that belong to you.<sup>1</sup> Any new datasets you add will belong to you.

The main page of the Dashboard will list all cruises that belong to you. 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. If there are mistakes, please notify [socat.support@noaa.gov](mailto: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.

All datasets are managed using their expocode, which is a unique<sup>2</sup> 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 code<sup>3</sup> 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.

---

<sup>1</sup> 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.

<sup>2</sup> 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.

<sup>3</sup> See <http://www.nodc.noaa.gov/General/NODC-Archive/platformlist.txt>

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.<sup>4</sup>

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, ...
```

Tags for metadata items are case insensitive. The tag is followed by either a colon or equals sign, which can have spaces around them. Tags for the expocode include 'expocode' and 'cruise expocode'. Tags for the ship/vessel name include 'ship', 'ship name', 'vessel', and 'vessel name'. Tags for the investigator names include 'investigator', 'investigators', 'investigator name', 'investigator names', 'PI', 'PIs', 'PI name', and 'PI names'. 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.

---

<sup>4</sup> Future versions of the Dashboard will require only the expocode to be given in the data file. The vessel (ship) name and the investigator name(s) provided are only used for display in the SOCAT QC Editor. Full metadata should be given in a separate metadata document at this time.

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 headers.<sup>5</sup> 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. The data columns and their suggested identities (their “data type”) will be shown to you in the “Identify Columns” page to allow you to modify these identities as needed. 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.

Once all columns are identified to your satisfaction, you ask the Dashboard to check your data for errors, using the “Check Data” button on the “Identify Columns” page. This will help verify that the columns are correctly identified. You will be notified of unreasonable values in the data, either because of incorrect data type, incorrect data units, or non-standard missing-data values. If errors are due to incorrect identification of the data columns, you can return to the “Identify Columns” page to correct the data column types and formats and recheck the data without having to upload the data file again. Note that all data samples must have a valid longitude, latitude, date<sup>6</sup>, and time of measurement. 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 work with such a dataset.

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

---

<sup>5</sup> Because of this, metadata lines can have blank entries (such as ‘,,,’ in CSV format files) following the metadata value.

<sup>6</sup> 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.

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.

After uploading the data for a dataset, you should examine plots of your data on the “Preview Data” page. These plots allow you to quickly detect inconsistencies or other errors in the data you provided. 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.

Before you submit a dataset for SOCAT, you must provide metadata for the data you have uploaded. 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. There are two ways of adding metadata to the Dashboard; at this time you can use either or both to satisfy the metadata requirement.

This first option is to provide metadata from the Online Metadata Editor (“OME”) for SOCAT<sup>7</sup> provided by CDIAC. You fill in this extensive form, which when completed is sufficient metadata for SOCAT, and save a copy (the “OME XML file”) to your local system. This file is then uploaded to the Dashboard using the “Edit OME Metadata” page.<sup>8</sup> When you provide this OME metadata, the date and time of the upload of this metadata will appear in the “OME Metadata” column in the Dashboard listing of this dataset. 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 CDIAC if you choose to do so.

The second option for providing metadata is to upload your own metadata file(s), with whatever content and in whatever format you chose, using the “Supplemental Documents” page. The files you upload will be simply 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

---

<sup>7</sup> See <http://mercury-ops2.ornl.gov/socatome/newForm.htm>

<sup>8</sup> Future version of the Dashboard will provide this metadata editor directly from the dashboard.

the new version of the document using the same name) or delete any supplemental documentation.

After uploading metadata for a dataset, and after verifying that the data appears correct in the preview plots of your data, you can submit the dataset to SOCAT using the “Submit for QC” page. Since all the data used for SOCAT must be archived, you are presented with three options for archiving the dataset you uploaded to SOCAT. 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. 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. 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.

When you submit your dataset to SOCAT, it will be added to the SOCAT QC Editor for SOCAT scientists to examine for quality control (“QC”) purposes. Once a dataset is submitted for QC, you are limited in what you can do with the dataset in 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<sup>9</sup> to “Suspended” and you can modify the dataset as desired. 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 place.<sup>10</sup>

---

<sup>9</sup> 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.

<sup>10</sup> 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.