Data Products Home
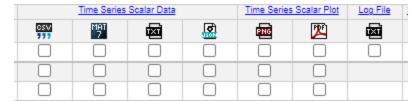
Data Products

Here's the place to learn and talk about our data products - both current and future. Data products are primarily available through Data Search (Data Search Help page), as well as the Data Preview page and Hydrophone Viewer tool.

We have also developed public web services / API to access these data products, which can be used to automate data requests from Oceans 3.0. The OpenAPI page is the best way to get started with the API, while our client-side code libraries make integrating the API into your code much easier (they're available via the Quick Start documentation). Once users have developed code or software using the API, they can then deploy it in The Oceans 3.0 Sandbox so that it runs local to the data on our cloud (contact us for more info). The API will also work with any cloud and high performance computing environment such as the Digital Research Alliance of Canada facilities. Running code, making use of the API and cloud options is particularly useful for automated, repetitive analysis and for handling large data sets.

The aim of this documentation is to help our users understand and make use of our data products. It is also useful as a detailed resource about data product specifications. This documentation is maintained as data products are improved and added. User input for new and improved data products is encouraged: you can email us (info@oceannetworks.ca) or press the request support button in the top right corner anywhere in Oceans 3.0.

This documentation is organized around the dataproductid with one page for each (these pages are direct child pages as one can see to the left and in the table below). The dataproductid (used internally) represents each data product group which have a common name, theme, purpose and usually have a number of formats contained within. For example, in the image below, the "Time Series Scalar Data" is the common scalar data product for parsed sensor data, such as temperature, with numerous formats available, each format representing the exact same data, just in a different presentation. Users can find direct links to the data product child pages in the Data Search Data Product Selection page. Simply click on the links that are also the data product headings and names, doing so will open a new tab with the specific data product documentation:

Users can also find links in the Data Search cart, in the Status column:

These links appear for most but not all data products (work in progress). Users can always find the data product specific documentation linked in the table below.

How does it all work? Need a good overall reference?

Please see our Oceans 3.0 article: https://doi.org/10.3389/fmars.2022.806452 This article details the complete data acquisition process, from the deep ocean to the data products users can see, interact with and download.

New Features and Highlights

- See the latest release notes: New Features Release Notes
- ONC Science Hub news stories on Oceans 3.0: check out the latest article here

Recently Updated

Downloading Data Products from Data Search

Current Data Products

ID	Data Product
1	Time Series Scalar Data (incl. stationary and mobile scalar devices)
2	Time Series Scalar Plot (incl. stationary and mobile scalar devices)
3	Borehole Temperature Time Series Plot
4	Log File
5	RDI ADCP Time Series
7	Audio Data
9	RDI ADCP Daily Current Plot
10	RDI ADCP Daily Intensity Plot
14	AVI Video
18	AGO Time Series Plot
19	BioSonics Time Series
20	Satlantic ISUS Time Series
21	Time Series Staircase Plot
22	Nortek Time Series (raw and processed formats)
23	MP4 Video
24	ASL Acoustic Profiler Time Series (AZFP, AWCP and ZAP echosounders)
25	CSEM Receiver Time Series
26	RDI Wave Time Series

Search Type / Sort By Options

Currently, the default search type to search and generate data products is "Instruments by Location". This is the option visible on Data Search Data Source Selection page (formerly known as Step 1), entitled "Sort By:". The default option, "Instruments by Location", allows user to navigate to a location and select an instrument/device category from which to request data products in Data Search Data Product Selection page (formerly know as Step 2). For these search requests, data from multiple instruments of the same category are combined together to form a continuous time series. For example, to form a long time series of CTD data in Saanich Inlet, an "Instruments by Location" stitches together the data from approximately 23 different CTD device deployments. The "Variables by Location" search type option is similar, except that users navigate to a location and select a sensor category variable, such as temperature, and then temperature sensor data is stitched together into a time series from the most appropriate devices at that location; also known as a "primary sensor search" or "search by water property". For many users, this search type is advantageous, as it does not require the user to know which device categories host which sensors and which of those are the best to use. For instance, many devices have temperature sensors, some of which are internal temperature for diagnostic purposes only, while the CTD (Conductivity Temperature Depth) devices generally have the best temperature sensors. The "Instruments By Category" search type option enables users to navigate to a device category, such as Hydrophone, then to a specific device (e.g. Ocean Sonics icListen HF Hydrophone 1252 (23155)) and then to a single device deployment (e.g. Cambridge Bay (03-Sep-2013 to 16-Sep-2014)). Users of this option are generally internal users and scientists who know the exact device they are interested in.

Data Product Options

For all scalar data products and many complex data products, users will be presented with options to customize their data products. This occurs after selecting a checkbox on the Data Search Data Product Selection page. These options are described in the individual data product pages. A compilation of the options is presented in the data product options page.

The term "scalar data product" refers to data from parsed sensors where there is one reading per time sample, often time series data or spatial data. These products are common and standard to all devices with sensors. They are also available at the device-level which is simply all the sensor data put together in a single file or plot. "Complex data products" are everything else, usually conglomerate device-level data products that contain all the data of the device such as raw or manufacturer format files and are usually specific to a device-category. These products are often more data dense, contain image or acoustic data.

Metadata

Metadata reports are available with nearly all different data products. These reports are produced automatically when a data search is completed and are made available via a link adjacent to the data, see step 3 in data search help. The reports contain extensive information about the data, including instrument location, deployment, calibration, data quality and data gaps.

Citation

The data products shall contain citation and attribution information wherever feasible, so that data products may be referenced and cited by persistent identifiers as described on the Data Citations page. All MATLAB MAT format file products contain a metadata structure with citation information, while the time series scalar text file products (CSV, JSON, ODV) have citation line(s) in their headers with the DOI URLs. Plot products have limited space and utility to include the full citation text, so a shortened version of the citation will appear below the plot body, listing the contributors, if organizations other than ONC contributed. All plots are capable of supporting a special logo, see the spectrogram data product as an example. For data products that cannot directly include the citation, such as binary data files, please refer to the Data Citations page for information on how to look up the citation and how to cite us. Future improvements will include better access to citation information (minting DOIs is a relatively new feature in our system).

Data Quality

Data quality information is supplied by way of data quality flags and comments in the data products, as well as annotations listed in the metadata reports. See the Quality Assurance Quality Control page for more information. This is well established for scalar sensor data, while complex data products may have specific quality flags or processes described on their corresponding documentation pages.

Data Availability

Data availability is indicated in step 2 of Data Search. The green data availability bar is based on archived data and may not show data for the last 24 hours (until it is archived). All data that goes through the shore-station drivers is archived in a raw format nightly as log files. Some devices provide data through FTP or HTTP file transfers; the data availability graph will be accurate in that case and data products will be available in near real-time (usually delayed by a few minutes). Although the data availability bar doesn't show it, scalar data is available live: data is usually only a few seconds delayed as it comes up the wire and through the various parsing, conversion, calibrated and QAQC steps. Many complex data products (data that is multidimensional, such as acoustic backscatter or profile data) produce data from log files. Since October 2015 these complex data products can access the raw data prior to archiving to produce near-live data, usually delayed by a few minutes. In all, users should be able to access near real-time data for all active devices, in addition to accessing historic data from as far back as 2002 (currently, we continue to acquire historic data).

Mobile Data

See the mobile device page to see how data products handle data from mobile devices. There are some spatially based data products, while most data products are time series.

Conventions

Time-stamps: Time-stamps are always in UTC. For file-names and string dates, the format conforms to the ISO8601 convention: yyyymmddTHHMMSS. In some cases, the millisecond portion may be added: yyyymmddTHHMMSS.FFFZ. Numerical time-stamps within data product files may follow a different format as noted on the data product pages. For instance, numeric time-stamps within MAT files are in the MATLAB serial date format. When resampling, the time-stamps are generally taken from the centre of the resample interval.

File-names: Note that the underscore character, "_", is used to separate the components of the names, while spaces, dots and other special characters are not included in file-names. File breaks are avoided as much as possible, but do occur for many reasons, including configuration or device changes, plus some data products have daily file breaks.

For an "Instrument by Category" search (see the "Sort By:" option in Step 1), files will be named as follows: DEVICECODE_SENSORNAME_yyyymmddTHHMMSS.FFFZ_yyyymmddTHHMMSS.FFFZ-MODE.EXT where:

- DEVICECODE is a descriptive string unique to each instrument.
- SENSORNAME is the sensor name as it appears in data search, and is only included if a single-sensor data product was requested.
- The first yyyymmddTHHMMSS.FFFZ is the time-stamp (ISO8601 format) of the first record in the file; the second yyyymmddTHHMMSS.FFFZ is the last time-stamp in the file (including data flagged and replaced with NaN). The date-to time stamp is not mandatory for all files; in particular, files streamed directly from the file archive will not get a data-to in their file-names. The time-stamp format optionally includes milliseconds: yyyymmddTHHMMSS.FFFZ, where 'FFF' are the milliseconds. If there is a data gap at the beginning and/or end of the search time range, the file-name dates will be different from the search time range. In the case of plotted products, the search time range sets the range of time axis so that users effectively control the horizontal scaling. Consequently, the file-name dates for plots will match the search time range, not the data time range.
- MODE is optional text which allows files of the same extension to be differentiated. It is used for different operation modes (Kongsberg scan or sweep for instance) or different data product options or multiple formats of the same extension. For example, scalar MAT files will get an 'ANCILLARY' when on ADCPs so they are not confused with RDI MAT files. Data product option mode strings are used on scalar data products primarily, examples: '-NaN', '-clean', '-

NaN_clean_avg15minute', '-MinMax1hour', see here for more details. Other data products supply file modes as described in their documentation.
- EXT is the file extension.

For an "Instrument by Location" search (see the "Sort By:" option in Step 1), files will be named as follows: STATIONNAME_DEVICECATEGORY_SENSORNAME_yyyymmddTHHMMSS.FFFZ_yyyymmddTHHMMSS.FFFZ-MODE.EXT, where:

- STATIONNAME is the station name, including node and station names separated by dashes, for example: *BarkleyCanyon-VPSUpperSlope.*
- DEVICECATEGORY is the device category, such as "CTD". If there is more than one device in the category, the file will contain multiple devices combined together for a long record of data.
- SENSORNAME is the sensor name and is omitted for a device-level data product that contains multiple sensors.
- yyyymmddTHHMMSS.FFFZ, MODE and EXT are as above.

For a "Variables by Location" search (see the "Sort By:" option in Step 1), files will be named as follows: STATIONNAME_variables_SENSORCATEGORY_yyyymmddTHHMMSS.FFFZ_yyyymmddTHHMMSS.FFFZ-MODE.EXT, where:

- STATIONNAME is the station name, including node and station names separated by dashes, for example: *BarkleyCanyon-VPSUpperSlope.*
- SENSORCATEGORY is the sensor category, such as "Temperature" or "Conductivity". If there is more than one device in the category, the file will contain multiple devices combined together for a long record of data.
- yyyymmddTHHMMSS.FFFZ, MODE and EXT are as above.

Data Search Size / Time Features

In Data Search, an estimated file size and processing time is given for each selected data product search in an open or completed cart. This value is based off of the size and processing time for the most recent previous search of the same format, options and approximate time range. A total estimated .zip file size is also given for the entire cart. The estimates provided are only accurate to an order of magnitude and are provided as a guide only. In cases where no similar searches have been run previously, the estimates will be "Undetermined".

Searches that will take longer than a week to run may be interrupted by software updates, while large file sizes (500 GB of data or more) may be difficult to download and manage on a local computer. In these situations, it is recommended to break up search requests into smaller time ranges, then downloading and processing that data before requesting more. This can be done programmatically via the Oceans 3.0 API / dataProductDelivery webservice. It is also recommended that users try a small search first (less than a day of time) and investigate / experiment before committing many large searches. Large searches may also be subject to interruption if system resources become taxed (this is usually only a problem for Hydrophone Audio Data). The size / time estimates are also used to trigger a pop-up warning to users. Users are also prevented from running too many searches at once, so that they do not block other users' requests. In that case, searches will be queued. All anonymous users share the same limit.

Email Notifications

Logged in users may choose to receive an email notification when their searches are complete. Also note that users may close or navigate away from Data Search and the searches in their cart will not be affected (this is also true for anonymous users if they allow cookies and always use the same browser on the same computer). If searches fail or are interrupted, ONC internal notifications are generated, alerting support staff who may contact affected logged-in users by email. Users will also receive emails when they use the Request Support button on the upper right side of all Oceans 3.0 pages.

Interoperability Partners

These inter-operable data products are no longer offered, but if they are of interest to you, please contact us.

ID	Data Product
15	ISDM Data Product
16	PANGAEA Data Product
17	POKM Data Product

130	McLane Sediment Trap Manufacturer Format
131	HoloSea 2D Image
134	Seismic Discrete Raw Data
136: 138	WERA Daily GIF data products
140: 144	136 137 138 (These are not likely to be used: 140 141 142 143 144)
139	Cast Scalar Multi-Profile Plot
146	Spectrogram For Hydrophone Viewer
148	CODAR Quality Controlled Surface Currents
150	CODAR Grid Files
147	Tiltmeter Engineering Data
152	Aligned and Depth-Binned Profile Data (On-Demand)
153	VEMCO Raw file
154	CODAR Data Availability Plots
155	Daily Collated Spectral Data For Archival Only
156	Weekly Collated Spectral Data For Archival Only
157	Standard Module Raw Data
158	Muon Tracker Raw Data
159	LiDAR Raw Data
160	PMT Spectrometer Raw Data
161	PMT Spectrometer TRB Counters
162	MXF Video
163	Mini Spectrometer Camera Rotated Color Corrected
164	Mini Spectrometer Camera Raw
165	Mini Spectrometer Image Data
167	Mini Spectrometer Sensor Data
168	Muon Tracker Sensor Data
171	PMT Spectrometer Sensor Data
172	LiDAR Sensor Data
173	CODAR Configurations
174	WOM Raw Data
175	WOM Sensor Data
176	WOM Log File
177	Digital Still

File formats

Additional resources for available file formats are available here.

Note that internal only formats may not be documented or listed here.

179	McLane Phytoplankton Sampler Data
180	McLane RAS Data
181	LIDAR Text File
190	AML Raw Data
191	AIS Raw Data
193	Community Fishers ISO 19115 XML Report
194	Magnetometer Raw Data
195	TiltMeter High-Rate Raw Data
196	Community Fishers CSV Annotations

If you have any data product related questions or would like to see additional data products, please let us know.