



## INTERNATIONAL YEAR OF THE SALMON 2022 PAN-PACIFIC HIGH SEAS EXPEDITION DATA POLICY

---

The 2022 International Year of the Salmon (IYS) Pan-Pacific Winter High Seas Expedition is a collaborative, international project to address the scientific hypotheses around salmon survival in the Pacific Ocean. The project's success, and its ultimate impact on science and society, relies upon professional coordination and data sharing across the project and the broader scientific community. A transparent Data Policy is essential to achieve the IYS science objectives, to facilitate collaboration, and deepen the impact of the IYS data. By participating in the International Year of the Salmon High Seas Expeditions you agree to the following Data Policy. The document is intended to be 'socially binding' with respect to the principles and common understanding described herein. The Policy does not preclude or replace any legal obligations or responsibilities that Participants or the institutions with which they are affiliated might incur.

### 1. Overview

The ultimate goal of this Data Policy is to facilitate data exchange between participating vessels and scientists and enable the publication of data from the International Year of the Salmon High Seas Expeditions to open-access domain-specific globally-integrated data repositories and license data for reuse under the [Creative Commons Attribution License](#). As well, we aim to facilitate the creation of an IYS Data Catalogue to act as a single point of entry to access all IYS related datasets. To facilitate these goals this Data Policy should build trust, clarify requirements, and establish a common foundation on which to build data exchange relationships.

The objectives of this Data Policy are to:

1. Provide clear expectations to expedition scientists and organizers regarding data access and release, as well as authorship and acknowledgment.
2. Ensure the fair and equitable use of IYS data and uphold the rights of individual scientists and institutions.
3. Encourage the rapid publication and dissemination of scientific data, results and knowledge, to support the involvement of a broad user community.

### 2. Definitions

- **IYS data:** Salmon ocean ecology data collected, measured, recorded, created or derived throughout the 2022 IYS Expeditions. This includes biological and biogeochemical data, derived data from analyzed

laboratory sample material collected at sea, sample and event metadata, satellite data products and data collected from Argo floats or autonomous glider(s).

- **Data Provider:** The Data Provider is defined as the person that owns and/or operates an instrument, creates and/or analyzes samples, produces a model output, or otherwise is in charge of producing a data set. The Data Provider is often the Principal Investigator of a particular research area and may be thought of as the data owner who has the authority to release the data under an open access license. The Data Provider is responsible for data quality and submission.
- **IYS Participants:** Participants of the Expedition whose scientific activities are officially endorsed by the IYS Steering Committee. Participants are bound to the IYS Data Policy and will have access to the processed IYS data before public release.
- **IYS Data Scientists:** Hakai Institute staff and the International Year of the Salmon / North Pacific Anadromous Fish Commission secretariat tasked with mobilizing data collected by the International Year of the Salmon Expeditions.
- **Raw data:** The data that are in the custom format of how they were collected and entered, or the format that is required by the scientists' home institution. It is data directly produced by sensors, devices, or manual observation, prior to additional processing, calibration and quality assessment/control (never modified).
- **Processed data:** Data that have had basic event and metadata quality assurance and result from data providers fitting their Raw Data into the [IYS Data Format](#).
- **Derived data:** Products based on raw data that may involve derivation of additional parameters or delayed-mode quality control using external data or post-use sensor calibration; model data or a combination with any external data, eg. by data assimilation, visualization, classification, or clustering.
- **Synthesis data:** Data from the IYS High Seas Expeditions combined with already published, historic or model data.
- **Data Packages:** A .zip file containing (for a [template Data Package see here](#)):
  - a statement of data accessibility: i.e. can the data be made public immediately, or at a later date. If at a later date, then justification is required.
  - the processed data (.xlsx or .csv format)
  - a data dictionary with definitions, units (SI units preferred), and a short description of each variable
  - a brief (300 word max) abstract describing what data were collected, how data were collected, and why data were collected
  - sampling protocols, calibration files and documentation of processing steps.

### 3. Metadata Standards and Supplementary Information

Standard metadata must be provided by Data Providers in order to populate metadata records in the [IYS Data Catalogue](#). These records conform to the Canadian Integrated Ocean Observing metadata profile, which is an extension of ISO 19115-2:2019 and ISO 19157:2013 metadata profiles. Required fields that must be provided by Data Providers of each Data Package include: names, roles and email addresses of dataset authors, dataset title, keywords, sampling instrument names, sampling depths, and limitations to interpretation of the data for external users. This information must be entered for each expected dataset via the [IYS 2022 Metadata Spreadsheet](#) (Note you may first have to request edit access to this document).

Supplementary materials must be submitted with a Data Package to provide additional contextual information about measurement details, methods, relevance, lineage, quality, usage and access restrictions of the data. Required supplementary materials are explained in the definition of a Data Package.

The [IYS Metadata Catalogue](#) will provide a centralized point of access and reliable, long-term storage for the metadata records for each individual data package produced. Metadata records will be stored during and beyond the duration of the IYS project. Metadata entered in the IYS Metadata Spreadsheet by Data Providers is entered into the IYS Metadata Catalogue ahead of the Expedition by the IYS Data Scientists or IYS Coordinators, and will be publicly visible. Once IYS data are processed, standardized and published to the relevant data repository, the metadata records in the IYS-OOS will be updated to include a link to the processed and standardized data. The responsibility for the accuracy of the metadata record lies with the Data Provider.

#### 4. Data Ingestion, Archiving, and Sharing among IYS Participants

The [2022 International Year of the Salmon GitHub Repository](#) will archive Processed Data and serve as the master copy of each Processed Dataset to share internally among IYS Participants. Only IYS Participants and IYS Data Scientists will have access to the data in this repository prior to publication. Data Packages containing Processed Data should be sent to the IYS Data Scientists ([secretariat@npafc.org](mailto:secretariat@npafc.org)) according to the timelines defined in this document (Table 1) for ingestion into the data archive.

Table 1. Description of IYS data categories and their schedule of availability to IYS participants (via 2022 IYS GitHub Repository) and public release.

Data Category	Example datasets	Availability	
		To Participants	To Public
At sea observations / Data reported in Cruise Report	Fish trawl, specimen measurements, bridge log, wildlife observations, CTD	Immediate after basic QA/QC	June 1st, 2022
Sensor data	Argo float / glider, CTD	Immediate after basic QA/QC	November 1st, 2022
Post expedition laboratory sample analysis	Stock ID, pathogen, physical oceanography	As soon as available	No later than March 31st, 2024
Satellite	Chlorophyll a, radiance	As soon as available	No later than March 31st, 2024
Metadata	All datasets	Before cruise	Before cruise

Early access by the IYS Participants to the data is crucial for successful collaboration. Hence, all processed data must be made available to all IYS Participants as fast as possible. Data included in the Cruise report following the Expedition should become publicly available at the same time as the report is published. Processed data shared

with the IYS Data Scientists will be stored on the 2022 IYS GitHub repository, and become freely available to all IYS Participants.

For IYS Participants who would like to use (analyze and publish) data collected during the 2022 IYS Expedition before data are released publicly, the Data Provider must be informed and offered collaboration on the scientific analysis and must be offered co-authorship based on the principles described in section “Authorship and Acknowledgement” below. The Data Provider may object to the usage of data in a publication if that publication conflicts with his or her own publication strategy. Any such objection must be discussed and agreed upon in writing with the IYS High Seas Coordinator and the Chief Scientist. The Data Provider may not object to the usage of data beyond the public release date.

For raw data archived on a national or institutional data storage platform, IYS Participants are encouraged to connect directly with the Data Provider to request access or ask questions. Raw data will not be stored on the 2022 IYS GitHub repository. The Data Provider is responsible for ensuring raw data longevity.

Exceptions to the processed data delivery deadlines can be made but must be discussed with the Chief Scientist (Evgeny Pakhomov) and the IYS High Seas Expedition Coordinator ahead of the Expedition and dispute resolution regarding unauthorized data use should also be taken up with the Chief Scientist.

## **5. Public Release of Data**

Good progress of a highly collaborative and interdisciplinary project like the IYS requires open availability of data to a wide user audience as early as possible. At the same time, it is important to acknowledge the substantial work that goes into collecting, quality controlling, formatting, documenting, and releasing scientific data. IYS policies pertaining to data use and acknowledgment aim to balance these two principles. Data access and usage policies change over time according to a staged process outlined in Table 1.

All processed IYS data will by default become freely and publicly available on the 2022 IYS GitHub repository based on the data type and data identified in Table 1. This table identifies the dates on which the GitHub repository will either be made public, or folders will be transferred to a public repository. Additionally, the metadata records in the IYS Metadata Catalogue will be updated to include a link to the publicly visible processed data. From this date on, there will be no restrictions on data usage aside from [Creative Commons Attribution licensing](#), but data users are strongly encouraged to communicate with Data Providers during early stages of all scientific analyses to ensure accurate usage and interpretation of data.

## **6. Data Standardization and Publication**

Clear, consistent documentation and standardization of IYS data will help to support a strong and lasting IYS data legacy, promote the broad and appropriate use of IYS data including the reuse and citation of data. The publication and mobilization of IYS data to open-access domain-specific globally-integrated data repositories will be supported by the IYS Data Scientists. To this end, IYS Participants will be provided an IYS Data Template to record core observations common between each vessel including Trawl Data, Specimen Data, CTD data, and Rosette Data. Data Providers should transform their Raw Data into Processed Data using the IYS Data Template and submit their data in this format.

For datasets other than those noted above, IYS participants are provided a general template for formatting using consistent station names, date formats, and location formats in the [data package example](#).

The ultimate goals for data publication are to ensure IYS data interoperability at both the project and global scales so that IYS data can be integrated internally and with other global datasets.

## 7. Authorship and Acknowledgment

Generally, **co-authorship** on publications and other public documentation must be offered to those that have **made a substantial contribution** following the [Code of Good Scientific Practice](#). An inclusive co-authorship approach is encouraged. Chief Scientists should be notified about planned publications that focus on data or samples collected at sea and should be considered for co-authorship if appropriate as defined by the Code of Good Scientific Practice. Co-authorship on publications and other public documentation must generally be offered to those that a) have made a substantial contribution to the creative process, that is, to the conception and design of the study, or to the analysis and interpretation of the data; b) have contributed to the preparation of the communications, reports, or publications that have arisen; c) be able to present in detail their contribution to the project and to discuss the main aspects of the overall research. Lead authors have the ultimate decision authority and responsibility to identify and appropriately engage co-authors.

IYS data will serve as a basis for derived or synthesis data products and manuscripts. To track the impact of IYS Data, and to comply with the Creative Commons Attribution License, IYS data must be referenced in publications and other public documentation, specifically by including relevant **digital object identifiers (DOIs)** and recommended citation found in the [IYS Data Catalogue](#) for each dataset.