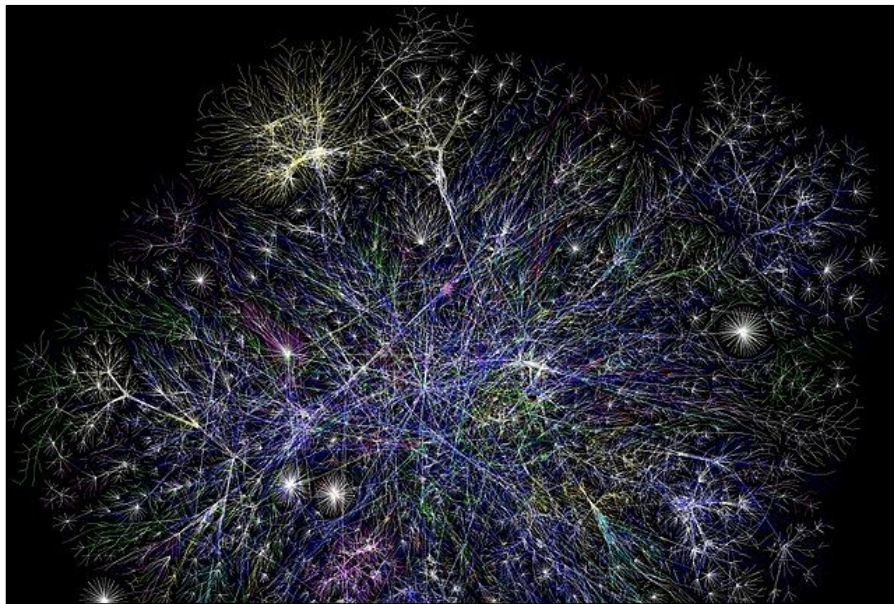


# International Year of the Salmon Data Mobilization



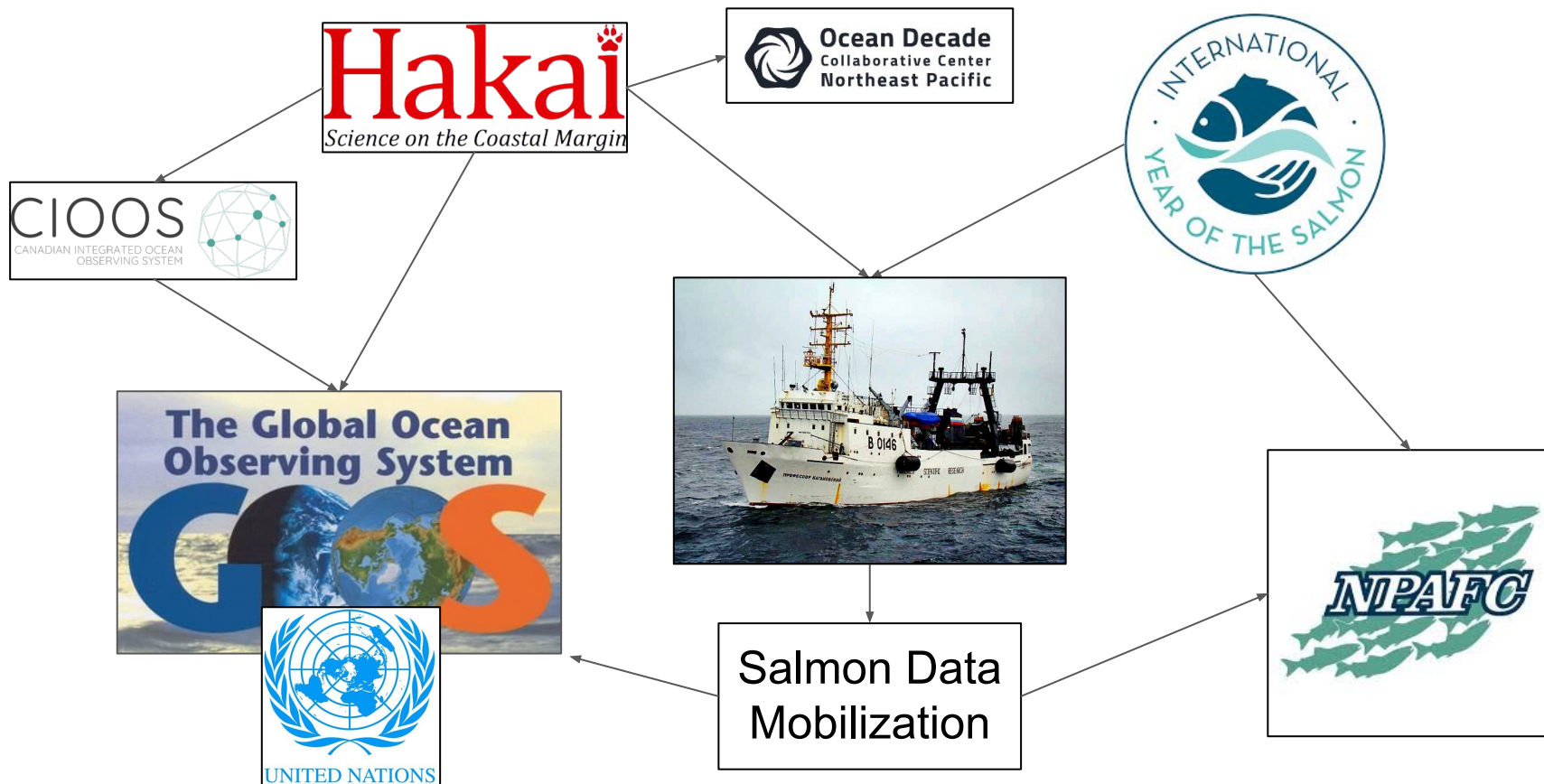
Brett Johnson, Tim van der Stap

# PRESENTATION OVERVIEW

---

- Big idea behind Data Mobilization
- Data Policy
- What services we've provided and what infrastructure we've built
- How you can contribute and access data
- Next steps
- Feedback

# BRIEF BACKGROUND



Data Mobilization: The process of making data available for appropriate re-use

- Salmon cross borders, so does our research
- Need new data methods to scale up
- Need data suitable for software
- Need institutional support





## 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.

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.

**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.

## 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.

# WHAT HAVE WE DONE AND BUILT?

## Services:

- Developed Strategic Plan
- Developed a Data Policy
- Built a Data Template
- Collected and curated metadata
- Standardized and published data for **global integration/access and project level integration and internal access**

## Infrastructure:

- **Data Catalogue**
- Data Hosting
- **IYS Data Mobilization Portal**
- Federated Metadata Architecture

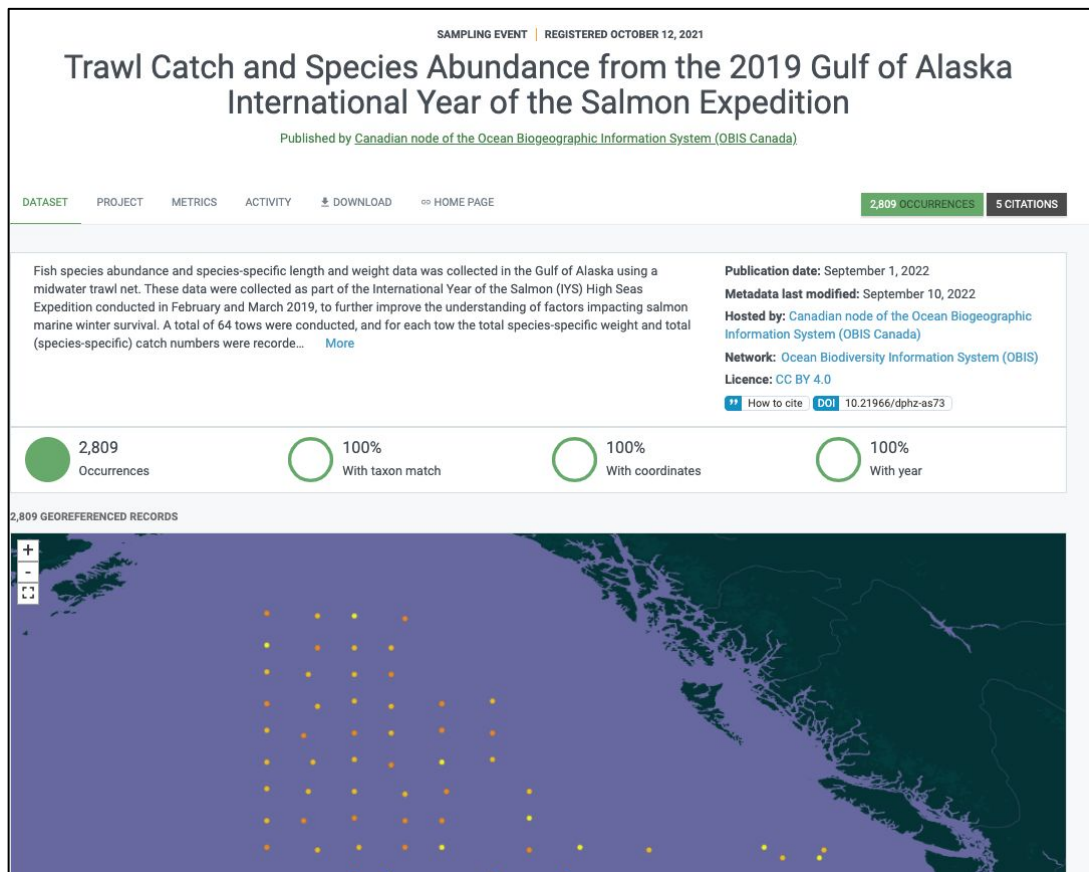


IYS Data Catalogue: <http://iys.hakai.org>

Centralized access to IYS Data

The screenshot displays the IYS Data Catalogue website. At the top left is the logo for the International Year of the Salmon. A navigation bar includes links for Home, Salmon & People, Get Involved, Events, Projects, About Us, Contact, High Seas Expeditions, and a highlighted DATA CATALOGUE link. Below the navigation bar, the breadcrumb trail shows 'Catalogue / Datasets'. On the left side, there is a sidebar with a map of Canada and the United States, a 'Filter by location' dropdown, a 'Year' filter with input fields and a 'Show blank range' checkbox, and a list of 'Ocean Variables' including 'Fish abundance and ...' (with 23 items) and 'Other' (with 22 items). The main content area features a '+ ADD DATASET' button, a search bar with the placeholder 'Search datasets...', and a search icon. It reports '65 datasets found' and allows sorting by 'Relevance'. Two dataset entries are visible: 'Satellite Chlorophyll-a Data from the CCGS Sir John Franklin collected during...' and 'Chlorophyll-a Data from the CCGS Sir John Franklin during the 2022 Internatio...'. Each entry includes a brief description and an 'HTML' link.

# HOW TO ACCESS DATA

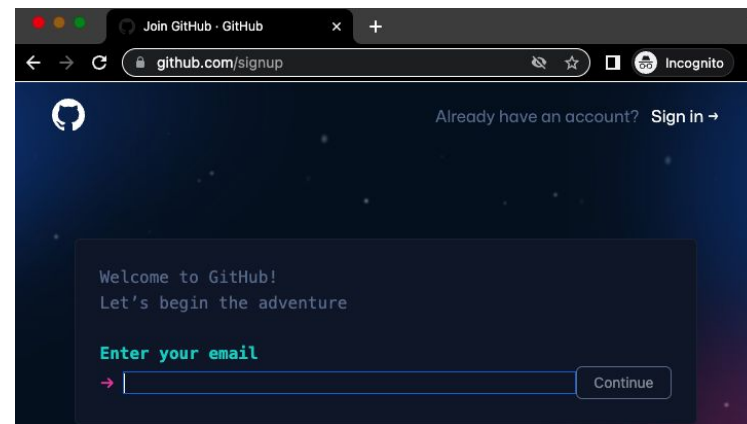


# To ACCESS INTERNAL DATA

Clicking on a link to internal data will result in 'Not Found'



Sign up for a GitHub account and send your username to [iys.data@hakai.org](mailto:iys.data@hakai.org) to be added to be granted access to internal repositories



## Google Datasets Search

- Google Datasets
- IYS Data Catalogue
- CIOOS
- DataONE
- OBIS
- GBIF

The screenshot displays the DataONE search results page. The top navigation bar includes 'Get DataONE Plus' and 'NEW'. The main header features the 'DataONE' logo and navigation links for 'Data', 'Services', and 'Community'. A red banner at the top of the search area says 'Clear all filters'. The search bar on the left contains the text 'International year of the salmon'. Below the search bar, the 'My Search' section shows the same text. The 'Filter by' section on the left has three expandable categories: 'Data attribute', 'Annotation', and 'Data files'. The main search results area shows 'Datasets 1 to 25 of 217'. The first result is 'Evgeny Pakhomov and Aleksey Somov. 2022. Salmon diet data from the 2020 Gulf of Alaska International Year of the Salmon Expedition.' with a DOI and SHA256 hash. The second result is 'Aleksey Somov and Evgeny Pakhomov. 2022. Trawl Data from the R/V TINRO during the 2022 International Year of the Salmon Pan-Pacific Winter High Seas Expedition.' with a DOI and SHA256 hash. A map on the right side of the page shows the search area with a grid overlay.

## IYS Data Mobilization Portal:


Centralized access to:

<https://international-year-of-the-salmon.github.io/about/>

- documents,
- data submission guidelines,
- data products,
- data publishing summaries

[IYS Data Mobilization Portal](#) [About](#) [Data Submission Guidelines](#) [Documents](#) [Data](#) [Tutorials](#)

## About



INTERNATIONAL  
YEAR OF THE SALMON

Welcome to the International Year of the Salmon Data Mobilization (IYS) Portal. This site is intended to primarily be an internal portal for IYS participants to centralize access to documentation, tutorials, data summaries, and data products.

Data Mobilization: The process of making data available for appropriate re-use.

**What is required by IYS scientists collecting data?**

1. Register the dataset you have collected by filling out [this metadata spreadsheet](#).
2. Submit standard data packages to the IYS ([secretariat@npafc.org](mailto:secretariat@npafc.org) and cc [iys.data@hakai.org](mailto:iys.data@hakai.org)) for internal sharing and publishing according to the [Data Submission Guidelines](#).
3. [Create a free GitHub account](#) and share your username with IYS Data Scientists by emailing us what your username is at [iys.data@hakai.org](mailto:iys.data@hakai.org) so that we can add you to the IYS organizational data hosting service and you can access and share internal data before it goes public (according to the [IYS policy](#)).
4. Keep track of where your samples are using the [Sample Tracking Spreadsheet](#)
5. Cite datasets used in publications according to the recommended citation for each specific dataset in the [IYS Data Catalogue](#)



# How To CONTRIBUTE YOUR DATA

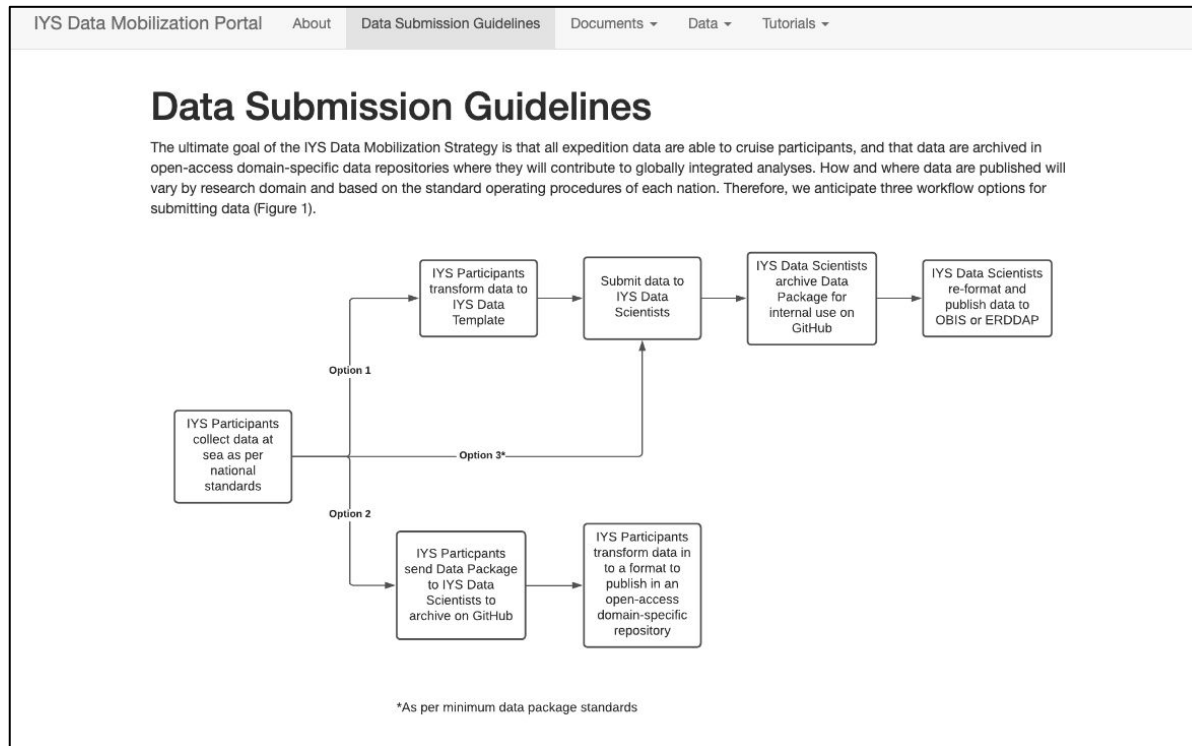
## Data Submission Guidelines:

### Data Package:

- Data files
  - IYS Data Template
- Data dictionary
- Readme file
- Supp. material

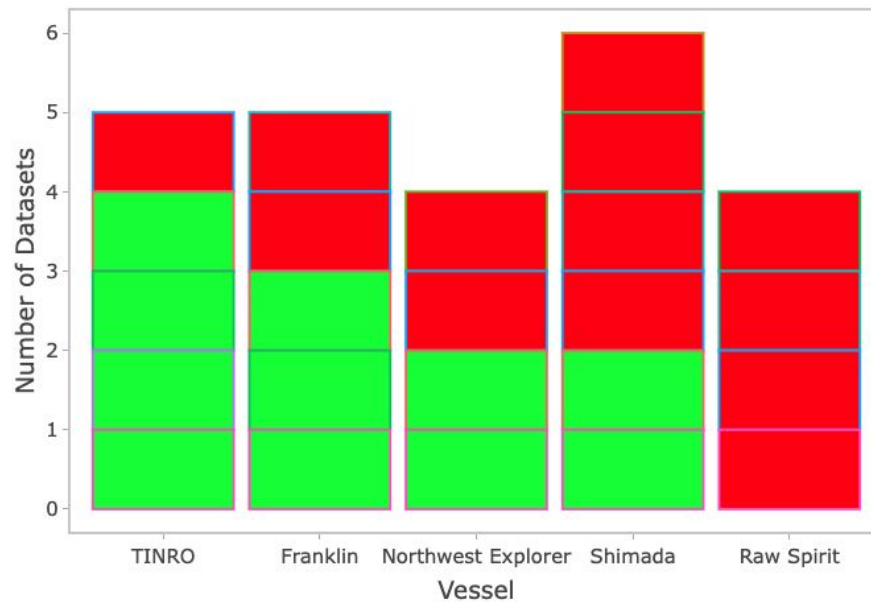
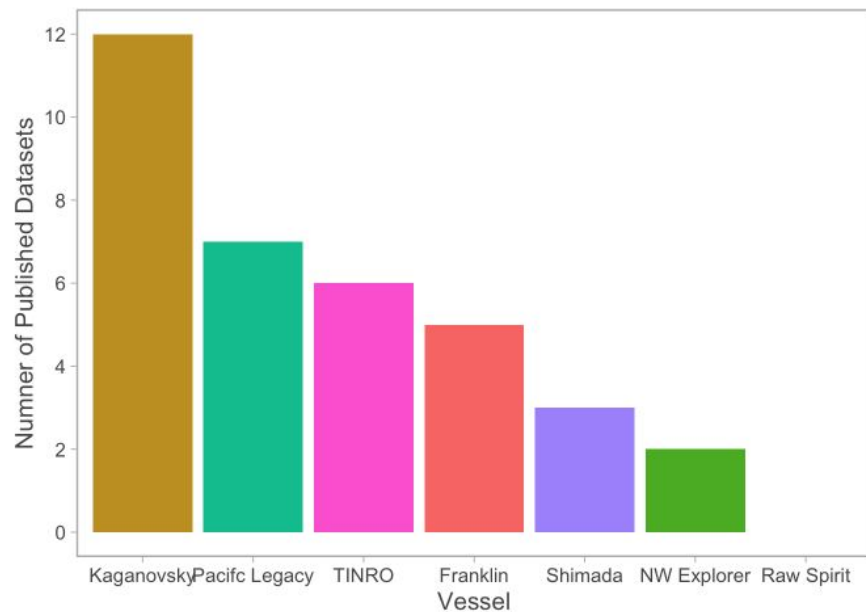
## IYS Data Mobilization Portal:

<https://international-year-of-the-salmon.github.io/about/>



# IYS DATA RECEIVED

Last updated: 2022-09-27



# IYS INTEGRATED DATA COLLECTION

## Integrating 2022

- Trawl,
- CTD,
- Biogeochemistry,
- Genetic Stock ID,
- Stomach Contents,
- Zooplankton

## Integrated Data Collection:

<https://doi.org/10.21966/81jn-6p63>

Integrated Data Collection from the International Year of the Salmon High Seas Expeditions

Responsible Organization

University of British Columbia

National Oceanic and Atmos...

Fisheries and Oceans Canada

Russian Federal Research In...

Hakai Institute

Pacific Salmon Foundation

Hokkaido National Research ...

National Oceanic and Atmosp...

Russian Federal Fisheries R...

North Pacific Anadromous Fi...

Laboratory of Fisheries Oce...

Pacific Branch of the Russi...

DatasetActivity Stream

MANAGE

Integrated Data Collection from the International Year of the Salmon High Seas Expeditions

This dataset is a collection of all of the International Year of the Salmon High Seas Expeditions trawl catch, collected specimen measurements, CTD and chemistry datasets that were collected at sea in 2019 aboard the Professor Kaganovsky, 2020 on the Pacific Legacy NO.1, and 2022 aboard the CCGS Sir John Franklin, NOAA Bell M Shimada, R/V TINRO, and R/V Northwest Explorer. These data span the entire North Pacific Ocean and were collected to better our understanding of the ocean conditions salmon experience during winter and how that may drive high seas salmon survival.

Access and Use

Licence: Creative Commons Attribution 4.0

Data and Resources

ZIP

Integrated IYS Data Collection (Internal...  
Collection of IYS datasets integrated into one excel workbook including all...

EXPLORE

PDF

Trawl Data from the R/V Bell M. Shimada during...  
Is a component part of the Integrated Collection

EXPLORE

PDF

Trawl Catch data from the 2019 Gulf of Alaska...  
Is a component part of the Integrated Collection

EXPLORE

- archived\_versions
- Changelog.txt
- figs
- input\_datasets
- IYS\_Integrated\_Data\_Collection\_V2.xlsx
- maps
- output\_datasets
- README.md
- scripts

# OUR NEXT STEPS

## Until March 31st:

- Expand Integrated Data Collection to include more data
- Take in new data and publish according to IYS Policy
- Continue publishing data to OBIS and ERDDAP via CIOOS and GOOS
- Continue publishing catalogue records
- **Not** publishing synthesis data sets from papers. Please publish those datasets as required by journals and cite the IYS Data Catalogue recommended citations for data used in your paper.

## Beyond March 31st:

- Transition to Open-Source project accepting volunteers
- Continue efforts to mobilize salmon data with partners and new projects including BECI

# THANK YOU



[Brett.Johnson@hakai.org](mailto:Brett.Johnson@hakai.org)

[Tim.vanderStap@hakai.org](mailto:Tim.vanderStap@hakai.org)

[iys.data@hakai.org](mailto:iys.data@hakai.org)



**Tula**  
TULA FOUNDATION





# DISCUSSION: CHALLENGES AND FEEDBACK

## Challenges for us:

1. Being brought on to the project after the first expedition
2. Unclear what our role should be
3. Communicating clear and simple expectations
4. Sheer volume of data and metadata

## Feedback:

Q. What role should organizations like the NPAFC, NASCO and PICES/ICES play in mobilizing data?

Q. Where should future data mobilization efforts focus more attention?

Q. What would make scientists more interested/less hesitant in sharing data openly?

Q. What services are missing from our Data Mobilization Infrastructure to make scientist's jobs easier?