

## A search and rescue team for salmon data

How the Data Task Force is increasing productivity for synthesis science

By Erin O'Reilly



A simple hook and line can be all one needs to catch salmon, but fishing for data about salmon is often more complicated. With a multitude of organizations collecting data all around the world, typically following differing protocols, the result can be a sea of data obstructed by a tangled mess of mismatched standards and different collection methods.

Fortunately there is an army of data scientists now focused on gathering, disentangling, and aligning data for other researchers called the Data Task Force. An NCEAS initiative launched in 2015 and funded by the [Gordon and Betty Moore Foundation](#), the Data Task Force rose from the realization that data collection, standardization, and management are often major constraints for big-data projects and can slow down researchers' ability to produce results.

“The Data Task Force is based on a simple research question,” said Jeanette Clark, data coordinator for the Data Task Force. “Can synthesis science can be done more efficiently if there is a team of people focused solely on the data needs of a project?”

So far, the answer is yes.

This team supports the [State of Alaskan Salmon and People](#) (SASAP), an initiative co-led by NCEAS and Alaska-based [Nautilus Impact Investing](#). The project is a perfect fit to test that research question, as it brings together multiple teams of researchers to assess the current state of salmon in Alaska from different ecological or social perspectives – analyses that depend on huge amounts of data.

Through in-depth search and rescue missions, the Task Force locates data from a wide variety of sources, which they then align to the same format, organize into datasets, and make them easily findable for future use. This helps eliminate the wild goose chase for data, an option usually unavailable due to budget constraints.

“The Data Task Force has changed the game of data,” says Ian Dutton, SASAP principal investigator and founder of Nautilus Impact Investing. “It has increased the efficiency of the

working groups, enabling more powerful synthesis and a comprehensive view of the entire system.”

In just two years, the Data Task Force has supported the creation of two incredibly large datasets. They’ve helped make sense of the tangled web of salmon fisheries data and provided scientists and managers with an easier way to fish for them: a streamlined line and hook.

As one line-and-hook example, the Data Task Force developed standardized datasets for SASAP teams and made them publicly available through the [Knowledge Network for Biocomplexity Repository](#), a repository for storing and sharing data that is co-administered by NCEAS.

The Task Force worked closely with the Alaska Department of Fish and Game to develop such a dataset for one SASAP team that is looking into why salmon sizes are declining and the associated consequences. To understand this, the researchers needed a seamless database of age, sex, and length for salmon across the entire state of Alaska over the last 50 years, a daunting task that may have been otherwise unattainable.



The mission of creating this database entailed gathering datasets from 13 sources, assigning geographic locations to these raw data, and reformatting them into a useable format. The result was a comprehensive and standardized dataset containing more than **12 million rows** of data that will enable the salmon researchers to answer their large-scale questions.

Creating such a database was unprecedented, requiring time, programming skills, and data management expertise that not every scientist has, explained Eric Palkovacs, lead researcher of the salmon-size team and professor at UC Santa Cruz.

“The Data Task Force allows teams to devote their time to developing interesting new ideas rather than struggling to assemble messy datasets,” said Palkovacs. “This greatly increases creativity and the scope of what data synthesis projects can accomplish.”

This expanded capacity enabled another SASAP team, which is focused on the governance of salmon use and conservation, to build an open-access database of Alaska Board of Fish proposals that stretches back to 1959, giving a record of stakeholder participation in fisheries management since statehood. By examining these proposals, managers can gain insights into the fisheries management needs of local communities from a previously unanalyzed source.

“This database pulls together the history of Alaska fisheries management from the voice of the people. It allows us to see how proposal topics have changed over time and what topics are especially important to local communities,” said Meagan Krupa, one of the Task Force’s data advisers. She helped develop a coding system for this database, which allows researchers to comb through and pull data from almost **100,000 pages of proposals**.

Since codes for proposals and other written materials may not always pick out the correct information, thereby threatening the data’s accuracy, the Task Force also verified their coding

methods, helping to make the process transparent and repeatable. Topping the effort off with technical training on building databases, the Task Force didn't just give the team their fishes but also taught them how to fish, so to speak.

"The idea of the Data Task Force excites me because I was able to learn more analytical methods for building a monster database," said Krupa.

According to Clark, the Task Force's data coordinator, their efforts help set SASAP apart from other working group models. They make inaccessible data easy to access, leading to faster results and less time spent on data disentanglement by the teams.

"It provides the working groups with a greater bandwidth to achieve high-level synthesis, a level yet to be performed for Alaskan salmon," she said.

*Header photo credit: Jason Ching; Photo of SASAP Salmon Size Declines research team: Palkovacs is back far left, Clark and Dutton are front right.*