

Guidelines on sharing data with SacPAS

SacPAS: Central Valley Prediction and Assessment of Salmon and other fishes, <https://www.cbr.washington.edu/sacramento/>

SacPAS Workshop, 6 December 2023

When collaborating with the SacPAS team (<https://www.cbr.washington.edu/sacramento/>), below are three checklists to help guide the process of: 1) sharing data; 2) changing, updating, and revising data; and 3) designating the file location. Additionally, to increase efficiency and reduce errors, it is generally a good idea to automate as many steps as possible. Automation decreases the potential for human-introduced errors during data handling and reduces people-hours spent on repetitive tasks that can be programmed and executed by machine. Still, the SacPAS team understands that there are many aspects to consider in long-term monitoring programs and for diverse data users; thus, there are tradeoffs in decisions of what and how processes can be automated. The SacPAS team strives to accommodate the needs of the data owners and bridge to existing processes.

1. Sharing data with SacPAS

The following will help create standardized, machine-readable data formats and automation.

- a) Keep data in a simple, consistent format that can be easily parsed by machine.
(e.g., tab, pipe, and comma delimited text files)
- b) Avoid complex file types, if possible. (e.g., .xlsx and .pdf files)
- c) Keep headers simple and contained on a single row.
- d) Include a single table in each file.
- e) Avoid using data delimiters within a data field.
- f) Clearly distinguish between null and zero values.
- g) Write comments in a separate column and not in place of a value.
- h) Describe the data quality as part of the data set.
(e.g., “provisional”, “preliminary”, “estimate”, “revised”, “approved” in a column)
- i) Keep data unaggregated to allow for comparison and integration with other data.
- j) Describe the methods for any summarized data or derived metrics.

2. Changing, updating, and revising data with SacPAS

Data changes, updates, and revisions should be included in the files.

- a) Include metadata in a file, separate from the data file, that is easily findable and accessible.
- b) Include all data as they accumulate through the season (or year) in the data file, rather than one data file for each subseasonal observation period.
- c) Announce or let the SacPAS team know when there are changes to the format.
- d) Keep the formatting of historical and current datasets equivalent or as comparable as possible.

3. Designating the file location

The following will help make the file location easily accessible by interested parties.

- a) Implement publicly accessible URLs or private/public ftp sites as the file location, if possible.
- b) Keep the file location consistent, if possible.
- c) Use filenames that are consistent and predictable.
- d) Announce or let the SacPAS team know if there are any changes to the file location.

Questions? Contact the SacPAS team at web@cbr.washington.edu.