

Data sharing: pathways across a range of efficiencies

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

SacPAS Workshop, 6 December 2023

The FAIR Principles help improve the efficiency and quality of shared data (FAIR: Findable, Accessible, Interoperable, and Reusable; <https://www.go-fair.org/fair-principles/>).


The diagram below is provided as an exercise to see how your data sharing pathway tracks from Findable & Accessible to Interoperable to Reusable, and where your pathway can track anywhere along the spectrum of efficient–inefficient data sharing.

Understandably, there are other considerations, limitations, and tradeoffs in how data are collected and shared, and that some datasets require more caution and discretion when sharing with the general public. Still, this exercise may provide some insight into where there could be greater efficiency in your data sharing workflow.

FAIR Principles	Efficient Data Sharing ←————→ Inefficient data Sharing			
Findable and Accessible	Data are published to website in a consistent location with a consistent naming scheme	Data are shared through personal communication or consistently to a location that requires approved access	Data are published to a website inconsistently or with inconsistent URL or filename	Data are kept in house
Interoperable	Data has consistent column headers in a consistent order		Data are inconsistent with column names, number of columns, or order of columns	
	Data types are consistent, or cells are left blank with a comment in the comments column.		Data types are inconsistent, comments are placed in integer rows, dates and times are not represented consistently	
Reusable	Data are the original collected data with corrections or omissions explained in a comments column	Data are aggregated by the day in a more detailed spreadsheet	Data are aggregated into a simple spreadsheet over an extended observation period	Data are aggregated down to a single number without clear methods
	Data are released in close to real time and later revised as data are corrected	Data are released in close to real time with no revisions	Data are published as a yearly revised dataset	Data are published over a year after collection

(For an example of a data sharing pathway, please see over.)

Example:

FAIR Principles	Efficient Data Sharing  Inefficient data Sharing			
Findable and Accessible	Data are published to website in a consistent location with a consistent naming scheme	Data are shared through personal communication or consistently to a location that requires approved access.	Data are published to a website inconsistently or with inconsistent URL or file name	Data are kept in house
Interoperable	Data has consistent column headers in a consistent order		Data are inconsistent with column names, number of columns, or order of columns	
	Data types are consistent, or cells are left blank with a comment in the comments column.		Data types are inconsistent, comments are placed in integer rows, dates and times are not represented consistently	
Reusable	Data are the original collected data with corrections or omissions explained in a comments column	Data are aggregated by the day in a more detailed spreadsheet	Data are aggregated into a simple spreadsheet over an extended measurement period	Data are aggregated down to a single number without clear methods
	Data are released in close to real time and later revised as data are corrected	Data are released in close to real time with no revisions	Data are published as a yearly revised dataset	Data are published over a year after collection