

EED SMARTT-2 Transition LDRD Proposal Template

Project Title: Seawater as a Strategic Resource I: Optimizing the MCRL data pipeline

Principal Investigator: Peter Regier

Project Team, if any: Tristen Myers Stewart

Bottom Line Up Front Statement: Seawater is MCRL's marine research portfolio depends on understanding Sequim Bay seawater chemistry. We seek to upgrade the existing MCRL data pipeline to allow real-time access, incorporate new sensing capabilities, standardize protocols, and improve scalability.

Science & Technology Challenge: PNNL's Marine and Coastal Research Laboratory (MCRL) is situated on Sequim Bay, which serves as a local testbed and seawater source for laboratories across campus. Projects spanning MCRL's project portfolio, including research on algal biofuels, aquaculture, biological monitoring, critical mineral extraction, ecosystem processes, and marine energy, all utilize Sequim Bay seawater. Because these research areas involve processes sensitive to changes in seawater chemistry, a comprehensive understanding of the chemical composition/variability of Sequim Bay seawater is crucial information.

To support this need, MCRL has made significant investments in monitoring infrastructure for marine chemistry and local environmental context (i.e., hydrology and meteorology), as well as digital infrastructure for cleaning, displaying, and distributing these data. EED's Directorate Objective for Ocean Solutions for Climate, Energy Security, and Coastal Resilience is supported through MCRL's monitoring suite, which provides relevant maritime data for research conducted in Sequim Bay. Additionally, this information is critical for ongoing development efforts, including MCRL's Research-accessible Shoreline (RASL), which will dramatically increase demand on our current data management/analysis capabilities.

The MCRLdata marine data pipeline provides access to data from a suite of instruments deployed off MCRL's dock. However, current data access is by request only, and new sensing capabilities brought online by specific projects are not currently available through the pipeline. Moreover, there is no central repository for laboratory datasets collected across projects relevant to Sequim Bay seawater chemistry, making it difficult to connect sensor-derived data to laboratory samples. **We have identified four opportunities to improve the accessibility, usability, and scalability of this data pipeline:** (Task 1) improving data intake; (Task 2) streamlining data upkeep; (Task 3) opening data access; (Task 4) codifying data training.

Project Description: This project will review and enhance PNNL's field data collection (PIs Myers, Hemery) and data management (PI Regier) capabilities related to the main value streams of seawater (aquaculture; energy generation, desalination, and mineral recovery).

Task 1: Improving Data Intake (Regier and Myers Stewart). This task will develop a modular field sensor and data management pathway for automating physical (Regier), geochemical (Myers), and biological (Hemery, Boise) data collection. This modular perspective will allow new users to rapidly integrate new sensing capabilities into the MCRLdata pipeline, including data processing.

Task 2: Streamlining Data Upkeep (Myers Stewart). This task will review existing capabilities and demand to generate an annual maintenance budget and protocol for field data collection. Sub-tasks will include (2.1) establish calibration and maintenance requirements and protocols for existing MCRL sensors. (2.2) Review and update existing MCRL data interfaces and protocols to better track usage, errors, deployment, and maintenance history.

Task 3: Opening Data Access (Regier). This task will build digital infrastructure for a real-time, on-demand data access pipeline for field and laboratory datasets, including visualization and statistical tools enabled by machine learning techniques. This repository will be access-controlled and all activities (downloads/code edits/etc) will be tracked to provide use metrics, and designed with scalability and modularity as cornerstones to enable easy assimilation of future data types. Subtasks will include: (3.1) Mirror the MCRLdata pipeline into a GitHub-based repository which will provide real-time, on-demand data access to staff researchers; (3.2) establish automated sensor health checks based on outcomes from Tasks 1 and 2; (3.3) develop statistical and data visualization tools, including including carbonate chemistry estimates, natural variability envelopes, and automated sensor error notifications. Machine learning will be crucial to this effort for anomaly detection in sensors, linking samples to sensor time-series, and understanding complex multivariate relationships.

Task 4: Codifying Data Training (Regier and Myers-Stewart). This task will create information and training materials to rapidly onboard seawater data users. We will design tools to be modular to facilitate future development, and include introductory materials, enabling others to develop project-specific tools independently.

Project Outcome(s): Improved access to a wider range of datasets will provide researchers with easier, more comprehensive access to relevant seawater data. Our proposed improvements will focus on integrating existing datasets (both sensors and samples), currently siloed across projects, into a single location. This will establish a clearer, more comprehensive view of our current data resources to identify existing knowledge gaps, provide a summary of historical Sequim Bay seawater conditions, and establish guidance for scaling up the number of data streams associated with RASL and other planned campus development efforts. Because GitHub is open-source and version-controlled, tools to visualize and analyze these data will be freely available to use and adapt by any user for project-specific needs. By developing standardized methods for bringing new data collection capabilities online, we will be able to provide consistent guidance to projects and lower the barrier to entry for integrating new data streams into the pipeline. We envision these improvements will enhance our capacity to collaborate across PNLL campuses with projects that could utilize these datasets, including projects across EED, EBSD, and NSD. Externally, this capability positions us to be responsive to current and future funding on seawater as a strategic resource, including DOE (e.g., WPTO, FECM, and Energy Frontier Research Centers), as well as industrial and state/federal agency partnerships (e.g., Washington Dept. of Health, DHS).

Funding Request: We request \$47,476,000 in salary for PI Peter Regier and project team Myers, Hemery, Boise, and Whiting to complete the tasks outlined above. A detailed breakdown of labor hours by staff member and task is described below. Project work will be completed by Sept 30, 2025 (end of Fiscal Year 2025) as required by the SMARTT program. Travel to Sequim during FY25 Q4 is covered for Regier (\$1950).

Task	Description	Gantt					Labor Hours						Travel	Task Total
		FY25Q3			FY25Q4		Regier	Myers	Hemery	Boise	Whiting	Cross		
		Apr	May	Jun	Jul	Aug	Sep	\$149.20	\$127.00	\$149.20	\$127.00	\$222.97	\$149.20	
1.1	Sensors							22	40	5	5	2		\$1,500.00
1.2	Budget							22	40	5	5	2		
1.3	Data Pipeline							80	20	5	5	2		
1.4	Training							16	40	5	5	2		
Total Hours								140	140	20	20	8	0	Total Request
														\$47,475.76