William Kumler

5515 15th Avenue NE, Apt #301

Seattle, WA 98105

909-907-2385 (personal mobile)

wkumler@uw.edu

May 4, 2025

Dear Schmidt Ocean Institute Hiring Committee,

I am applying for the Data Solutions Engineer position in the Science and Data Division of Schmidt Ocean Institute. SOI's position as the premier privately funded oceanographic research facility operator is a major factor in my desire to [work for you]. Your commitment to public sharing of data aligns with my personal passion of making both data and the ocean more accessible to the public for outreach, education, and inspiration purposes. I will be graduating as a Ph.D. in chemical oceanography with a specialization in data science in July 2025 and can offer a highly technical skill set focused on the collection, stewardship, and dissemination of oceanographic data.

I graduated from UC Berkeley in 2018 with degrees in marine science and molecular environmental biology and began working in both a biomechanics wet lab with Dr. Mimi Koehl and a mass spectrometry lab with Dr. Bethanie Edwards. I then joined the Marine Microbial Metabolomics Research Center under Dr. Anitra Ingalls at the University of Washington's School of Oceanography for my Ph.D. work. The past six years have resulted in multiple peer-reviewed publications, several technical projects that have been adopted by the wider scientific community, and nearly three months of total seagoing ship time.

While my recent focus has been on mass-spectrometry data, I was drawn to the field by the breadth of data available and the degree to which it was underutilized. My early thesis work consisted of refining existing peak-detection algorithms for the detection, identification, and quantification of molecules in the marine environment. This data-driven approach contrasted with the previously restricted searches that the lab and was only possible given my background in the R programming language. Frustration with the slow and convoluted methods available for rapid exploration and analysis of mass-spectrometry data resulted in my first publication, proposing a novel framework for data access that I then implemented as an R package and has since been adopted by Metabolomics Workbench and used by labs across the globe. I've continued to develop this framework and have recently translated it into Python and am in the process of implementing it as a database schema for use with DuckDB and Parquet, all freely available on GitHub.

Other thesis chapters used this framework to explore oceanographic data collected during the Simons Collaboration on Ocean Processes and Ecology (SCOPE). My participation in three research cruises, a month-long fieldwork incubation project, and my experience working with data from many additional ones highlighted that oceanographic data is often messy and benefits enormously from the metadata and access principles inherent in the FAIR guidelines. I often wished that the SCOPE project had funded a role exactly like SOI's current opening to streamline collaboration between labs and make data products accessible to both other researchers and the public through curation and visualization.

The SOI Data Solutions Engineer position is the perfect combination of my experience in oceanographic data analysis and my desire to facilitate future data science. I am deeply enthusiastic about the opportunity to steward data from collection on the Falkor (too) all the way through publication both informally for the general public via data-centric applications and rigorously for scientific publication and reuse. SOI's commitment to advancing the frontiers of ocean exploration and research is one that I look forward to enabling both professionally and personally.

Sincerely,

William Kumler