



## Exposing SAMOS Data and Vocabularies within the Semantic Web

Nkemdirim Dockery, Jocelyn Elya, and Shawn Smith

COAPS, Florida State University, Tallahassee, United States (nd08f@my.fsu.edu)

As part of the Ocean Data Interoperability Platform (ODIP), we at the Center for Ocean-Atmospheric Prediction Studies (COAPS) will present the development process for the exposure of quality-controlled data and core vocabularies managed by the Shipboard Automated Meteorological Oceanographic System (SAMOS) initiative using Semantic Web technologies.

Participants in the SAMOS initiative collect continuous navigational (position, course, heading, speed), meteorological (winds, pressure, temperature, humidity, radiation), and near-surface oceanographic (sea temperature, salinity) parameters while at sea. One-minute interval observations are packaged and transmitted back to COAPS via daily emails, where they undergo standardized formatting and quality control. The authors will present methods used to expose these daily datasets.

The Semantic Web, a vision of the World Wide Web Consortium, focuses on extending the principles of the web from connecting documents to connecting data. The creation of a web of Linked Data that can be used across different applications in a machine-readable way is the ultimate goal. The Resource Description Framework (RDF) is the standard language and format used in the Semantic Web. RDF pages may be queried using the SPARQL Protocol and RDF Query Language (SPARQL).

The authors will showcase the development of RDF resources that map SAMOS vocabularies to internationally served vocabularies such as those found in the Natural Environment Research Council (NERC) Vocabulary Server. Each individual SAMOS vocabulary term (data parameter and quality control flag) will be described in an RDF resource page. These RDF resources will define each SAMOS vocabulary term and provide a link to the mapped vocabulary term (or multiple terms) served externally. Along with enhanced retrieval by parameter, time, and location, we will be able to add additional parameters with the confidence that they follow an international standard.

The production of RDF resources that link daily SAMOS data to descriptors such as parameters, time and location information, quality assurance reports, and cruise tracks will also be described. The data is housed on a Thematic Real-time Environmental Distributed Data Services (THREDDS) data server, so these RDF resources will enable enhanced retrieval by any of the linked descriptors. We will showcase our collaboration with the Rolling Deck to Repository (R2R) program to develop SPARQL endpoints that distribute SAMOS content. R2R packages and transmits data on a per cruise basis, so an immediate result of the SAMOS exposure will be the narrowing of the gap between expedition type data (e.g. R2R cruises) and SAMOS observatory type data.

The authors will present the development of RDF resources that will collectively expose shipboard data, vocabularies, and quality assurance reports in an overall structure which will serve as the basis for a COAPS SPARQL endpoint, enabling easier programmatic access to SAMOS data.