Enabling Quality Assessment in the EarthCube Implementation

Janet Fredericks
Woods Hole Oceanographic Institution

This paper reflects my own views, which have evolved through participation in meetings and projects listed in the affiliations below.

Affiliations:
QARTOD (Quality Assurance in Real-Time Oceanographic Data);
Lead-PI in Q2O (QARTOD to OGC)
MMI (Marine Metadata Interoperability Project)
Institutional Member OGC
Member of the IOOC/DMAC-ST
Senior Member IEEE
NERACOOS DIF Member

Premise

There is great value in sharing data, for the data provider, as well as the research and observational community at-large. But, sharing can put too much demand on data managers with little support or guidance. To add value to and minimize the effort, data providers need to be encouraged to output data in a standards-based framework (such as OGC Observational Web Services) and large-programs should be encouraged to ingest data and metadata in standards-based frameworks. Tools need to be developed to make it easier for data providers to output common sensor-based earth observations in standards-based frameworks, along with fully described systems of how the observations were derived. Independent registries will enable sharing of common vocabularies and ontologies. Domain experts must be engaged to map relationships across disciplines and domains. Movements to engage the larger community are falling off, due to a lack of funding (MMI/QARTOD). This will leave behind the "dark-data" of those not involved in big programs (such as OOI/IOOS), which is most of the federally funded research. By working within each community but guiding to use a common approach, we can implement an EarthCube framework that promotes participation and well-described observations, enabling quality assessment. By investing in the ability to easily encode information about sensors, instrument configuration, meaningful processing descriptions and deployment history, our national programs can build earth observing systems with a solid foundation of understanding and trust.

Problems

I manage a coastal observatory off the coast of New England, which has power and fiber-optic Ethernet-enabled connectivity to shore and the WWW. The project was funded by NSF and became operational in 2001. In 2004, an NSF program manager came to Woods Hole to discuss our operations. She insisted that we develop interoperable solutions to data management, even though our scientific users were quite happy with our operations which fed data in real-time through text based summaries served by an ftp service and the JGOFs data management system. I easily installed some example services through the OOSTethys cookbooks, enabling data to be described and encoded in OGC SOS/SWE. But, there was no means of communicating sufficient information to assess data quality of the real-time data stream. And none of the federal agencies or programs ingested the OGC/SWE encoded data. But, it soon became obvious, that people were interested in these data. Each interested party requested specialized software to be installed and format changes in our output. Modelers want NetCDF and THREDDS servers; NDBC asked for format changes and specialized software to be installed; Google and RSS feeds were requested, each requiring yet another means of data delivery. OOI says to participate we "just have to install" JASON. Each format specification and software installation puts a drain on the data management system to provide services beyond the scope of the funded project. It also requires additional communication when a system changes (sensor swap out; reconfiguration; maintenance etc). This is not a scalable model for sharing data. Yet, sharing with each program not only benefits the programs, but can provide a benefit to the data provider. For example, the OOI program will provide additional integration and visualization tools for the MVCO user. Data for the development of better models will enhance our opertations.

SOLUTION

Data providers can provide data to meet the needs of its users and data can be offered to the world through standards-based encodings. If these encodings sufficiently describe the sensor capabilities, configuration, processing and deployment history, a solid foundation in sharing observations for purposes beyond the project goals is enabled.

The Q2O project developed a model implementation demonstrating how to encode a fully-described observing system in OGC SensorML and the observations in O&M. Included in the model is a means of associating meaning (imparting knowledge) for QC flags, processing for derived parameters, event history and provenance. The model includes two simple concepts:

- integration of resolvable links in the process descriptions (Figure 1)
- role-based structuring of SensorML enabling the development of tools and encodings by the responsible parties (Figure 2)

Resolvable links were created using the MMI Ontology Registry & Repository (ORR). Each term (processes, platforms, observable properties, observed properties, qcflags) is fully-described, with definitions, equations, references, etc. By registering the vocabularies with the ORR, the vocabulary is encoded in Ontology Web Language (OWL), providing a framework for creating relationships in a standards-based framework (RDF).

Role-based structuring of SensorML encoding will enable adoption in stages of fully-described systems. If we first engage sensor manufacturers, we can supplement human-readable manuals with machine-harvestable descriptions of important sensor capabilities relating to quality assessment. The approach should span domains, but the implementation should occur through communities working with domain experts. Configuration applications would output relevant SensorML that describes information in the set-up that relates to quality. Then tools would be needed to integrate the files and supplement them with further QC and processing information.

Use Cases

NOAA/NDBC has recently agreed to ingest the OGC/SWE waves SOS that we offer at the Marthas' Vineyard Coastal Observatory (http://q2o.whoi.edu/node/129). In our implementation, QC flags have definitions with meaning and links to the tests that generated each QC flag. Each test is also well-described with a resolvable link (URL). Our QC tests can be mapped to other agencies' QC tests with the same meaning (mvco/spikeTest=whoevers/rmOutlier). Our QC flag meaning can be mapped to another QC flag meaning, even though it is called something different (IntensityFlag=SignalStrengthFlag) and have different values (pass=0 or pass=3). Even with a range check failure (I really didn't expect 5 meter waves), the data was not filtered out and NDBC sees the large waves from TS Irene and because the QC flag was set, its 24/7 operators are alerted to the event.

MVCO provides meaningful information about the methods of computing waves from particle motion in the water. In the descriptions, one might learn the limits of its use for repurposing. What is the frequency of computation of water velocity? Is it fast enough to observe wind waves? Is the sampling period long enough to see long-period swell? Although we have configured the sensor to meet our objectives, other applications may have different requirements. There is more to sharing an observation that the observation itself. One must be able to clearly understand how the observed property became the observation. This can be done through documenting the sensor, processing and deployment information in a standards-based encoding and serving it through a standards-based framework.

Figure 1. Well-described sensors, configurations, deployments, processing and provenance in standards-based encodings (OGC) enables machine-to-machine harvesting of information that enables dynamic quality assessment. By referencing standards based vocabularies (OWL/RDF), relationships across-domains can be built, promoting interdisciplinary sciences on a global scale.

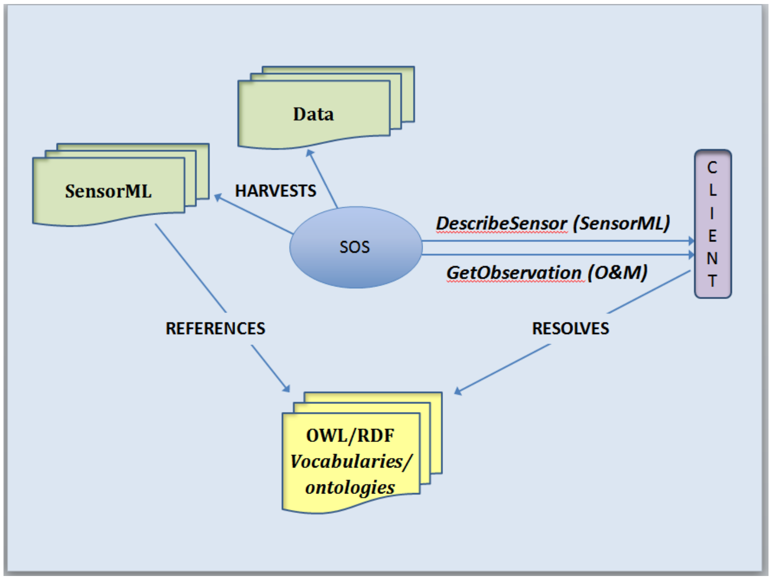
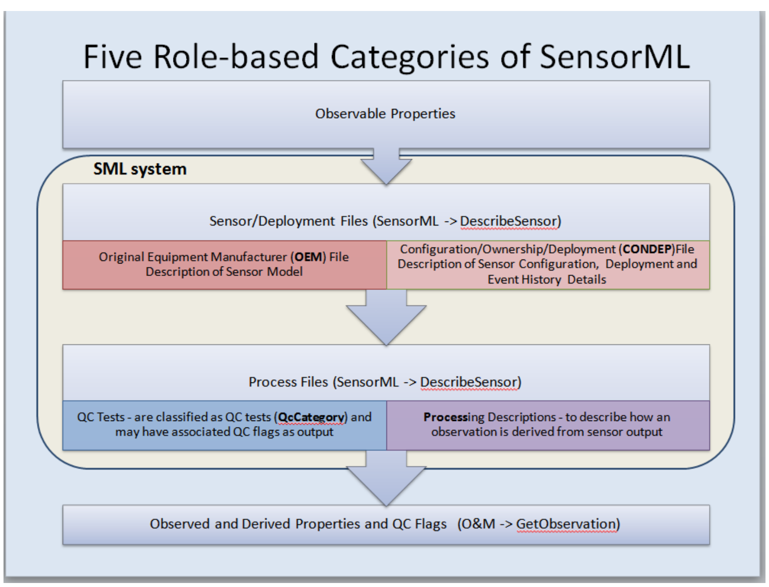


Figure 2. By adapting a structured approach, we can begin the slow path to adoption by encouraging manufacturers of common earth-based observing sensors to encode important information about sensors in SensorML. This will enable knowledge of instrument precision, accuracy, performance curves, configuration etc to be communicated in a machine-to-machine harvestable framework, instead of buried in a manual. In parallel, groups can develop and share well-described QC algorithms, which can be referenced in the SensorML descriptions. Also, means of deriving properties from sensed properties can be documented with authoritative references.



Bibliography

Fredericks, Janet J. and Carlos Rueda, "Content-Infused OGC Web Services Enabling Dynamic Quality Assessment in Observing Systems", SWE 2011 Workshop, Calgary, AB. 2pp. (in press)

Fredericks, Janet J., Mike Botts, Luis Bermudez, Julie Bosch, Philip Bogden, Eric Bridger, Tony Cook, Eric Delory, John Graybeal, Sara Haines, Anne Holford, Carlos Rueda, Jordi Sorribas, Barry Feng, Christoph Waldmann, "Integrating Standards in Data QA/QC into OpenGeospatial Consortium Sensor Web Enablement ", in *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2),* Venice, Italy, 21-25 September 2009, Hall, J., Harrison D.E. & Stammer, D., Eds., ESA Publication CWP-4C-04.

Ruedas, Carlos, **Janet Fredericks**, Luis Bermudes, "The MMI Ontology Registry and Repository: A Portal for Marine Metadata Interoperability", IEEE Xplore OCEANS/MTS 2009, 6pp.

Bosch, Julie, **Janet Fredericks**, Mike Botts, Tony Cook, Sara Haines, Philip Bogden, Eric Bridger, "Applying Open Geospatial Consortium's Sensor Web Enablement to Address Real-Time Oceanographic Data Quality, Secondary Data Use and Long-Term Preservation", IEEE Xplore OCEANS/MTS 2009, 5pp.

Fredericks, Janet J., Mike Botts, Julie Bosch, "Integrating Standards in Data QA/QC into OpenGeospatial Consortium Sensor Observation Services ", IEEE Xplore OCEANS/MTS 2009, 10pp.