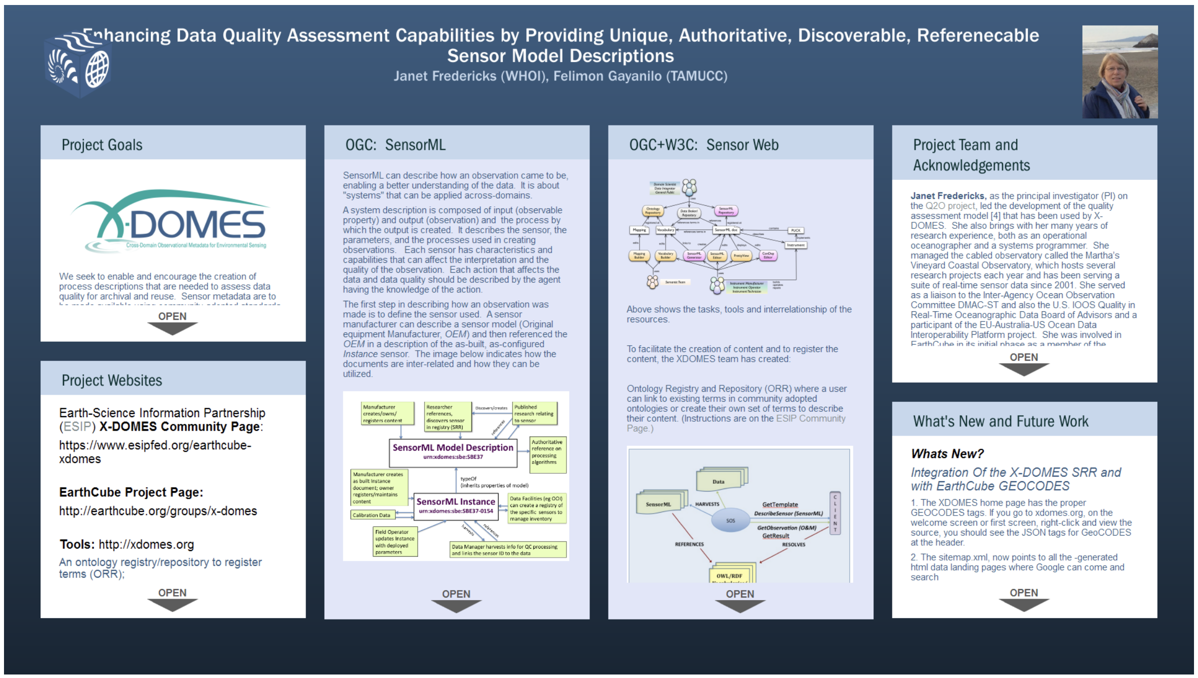
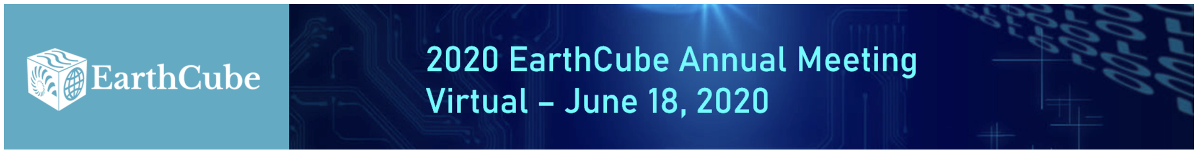
Enhancing Data Quality Assessment Capabilities by Providing Unique, Authoritative, Discoverable, Referenecable Sensor Model Descriptions



Janet Fredericks (WHOI), Felimon Gayanilo (TAMUCC)

PRESENTED AT:



PROJECT GOALS

X-DOMES

Cross-Domain Observational Metadata for Environmental Sensing

We seek to enable and encourage the creation of process descriptions that are needed to assess data quality for archival and reuse. Sensor metadata are to be made available using community-adopted standards (OGC/W3C) to assure FAIR data practices in our earth observations and its associated metadata.

The team has created tools for the non-expert to create and manage SensorML documents. The tools enable the creation of machine-actionable content.

X-DOMES implementation will:

- Help large observational data producers automate and manage sensor and operational provenance
- Encourage small federally funded data providers to describe sensor data in ways that meet agency requirements for data management
- Facilitate creation of and access to common content using standards-based production of interoperable sensor documentation by environmental sensor manufacturers
- Enable data aggregation centers to build relationships across domains for integration of sensor-based observations
- Provide the ability to assess data quality and automate quality control, based upon manufacturers' descriptions of sensor provenance
- Generate registries for sensor and deployment metadata that can be utilized by building blocks of a layered architectural cyberinfrastructure [1]

- Create open-source tools and libraries to discover, access, translate and associate sensor metadata, providing unique identifiers for sensor models, as-builts and processes.
- Promote better documentation for data archival of federally funded assets
- Speed sensor network deployments, providing better, faster event response data
- Reduce data analysis effort and time for scientists and emergency managers
- Improve the reproducibility of research products by capturing relevant metadata at each stage of data generation

PROJECT WEBSITES

Earth-Science Information Partnership (ESIP) **X-DOMES Community Page**: (https://www.esipfed.org)

https://www.esipfed.org/earthcube-xdomes (https://www.esipfed.org/earthcube-xdomes)

EarthCube Project Page: (http://earthcube.org/groups/x-domes)

http://earthcube.org/groups/x-domes (http://earthcube.org/groups/x-domes)

Tools: http://xdomes.org (http://xdomes.org)

An ontology registry/repository to register terms (ORR);

A Basic SensorML creator/viewer/editor

An Advanced SensorML creator/viewer/editor

A SensorML registry/repository (SRR) to register SensorML documents

Standards:

http://www.ogc.org/standards/sensorml (http://www.ogc.org/standards/sensorml)

https://www.w3c.org/standards/semanticweb/ (https://www.w3c.org/standards/semanticweb/)

https://www.w3.org/TR/vocab-ssn/ (https://www.w3.org/TR/vocab-ssn/)

(specifically the alignment of the OGC SensorML with the W3C SSN/SOSA

Semantic Web Guidance:

http://marinemetadata.org (http://marinemetadata.org)

GitHub Repository

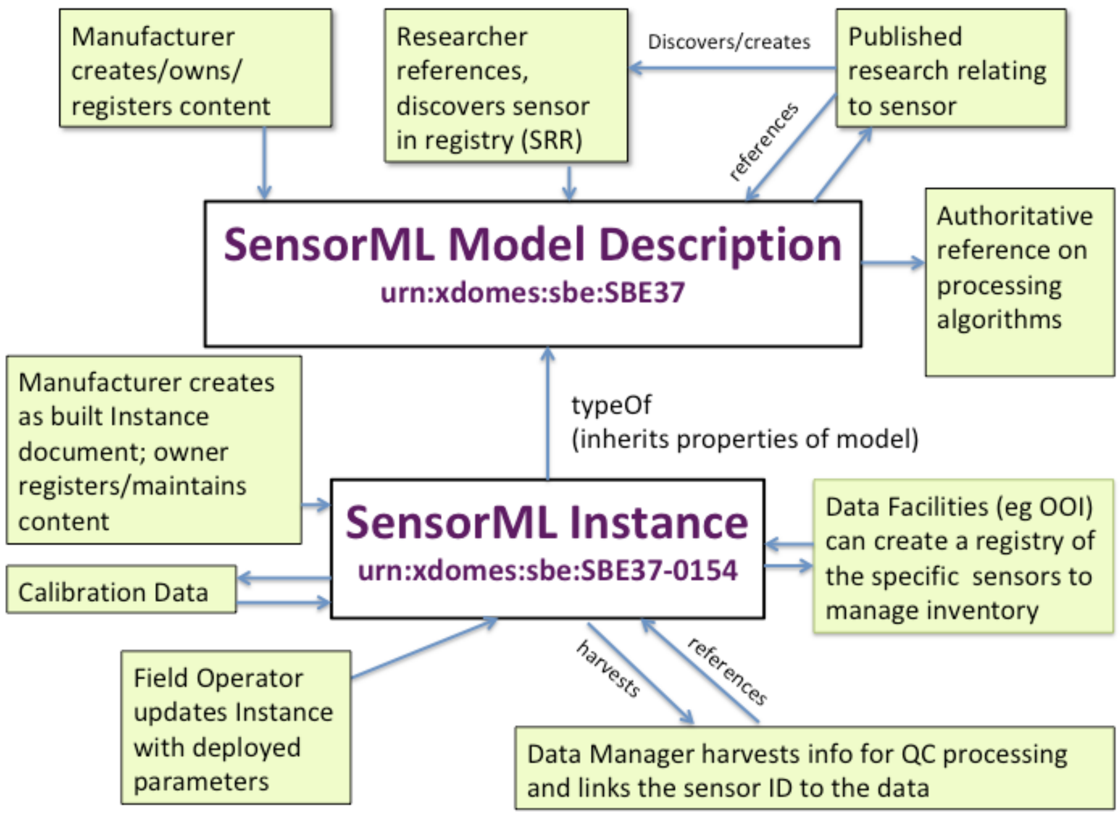
https://github.com/xdomes (https://github.com/xdomes)

OGC: SENSORML

SensorML can describe how an observation came to be, enabling a better understanding of the data. It is about "systems" that can be applied across-domains.

A system description is composed of input (observable property) and output (observation) and the process by which the output is created. It describes the sensor, the parameters, and the processes used in creating observations. Each sensor has characteristics and capabilities that can affect the interpretation and the quality of the observation. Each action that affects the data and data quality should be described by the agent having the knowledge of the action.

The first step in describing how an observation was made is to define the sensor used. A sensor manufacturer can describe a sensor model (Original equipment Manufacturer, *OEM*) and then referenced the *OEM* in a description of the as-built, as-configured *Instance* sensor. The image below indicates how the documents are inter-related and how they can be utilized.



Our team has created two versions of an online SensorML editor/viewer.

- The BASIC editor is easier to navigate but limited in its ability to incorporate full descriptions and links to terms.

- The full-featured editor requires some functional knowledge of how to describe your terms. There is a video available on the ESIP Community Page (https://www.esipfed.org/earthcube-xdomes).

.

Once the content is created. It can be registered and maintained by the XDOMES SensorML Registry and Repository (SRR). This act provides unique identifiers that can be used in associating the sensor to data and publications.

The tools are ready for stakeholders to begin creating and registering the descriptions of sensor models. And to describe a sensor Instance (as built sensor) that provides a unique description of a particular sensor. These documents can be registered - thereby creating a unique ID that can be referenced in data systems and publications.

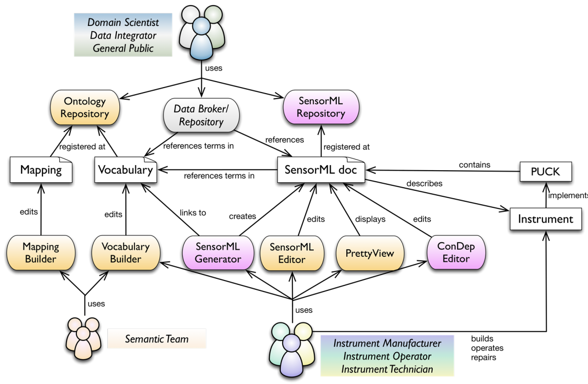
TO GET STARTED: Check out and save a file from the SRR, then load it into the Basic Editor. Then start describing your sensor!

SensorML Contents

- Description/Header (gml:urn, name, keywords)
- Identifiers (sml:name, model number)
- System Classifiers (type of sensor, etc.)
- Characteristics (weight/length etc)
- Capabilities (precision/accuracy info relating to data quality)
- Contacts
- Documentation
- Input (observable properties)
- Output (data descriptions)
- Parameters (used to create output from input)
- Components (eg., conductivity cell, temperature sensor, pressure sensor)
- Modes (versions of the sensor model)

* the greyed out categories are not in the basic editor as of June 2020.

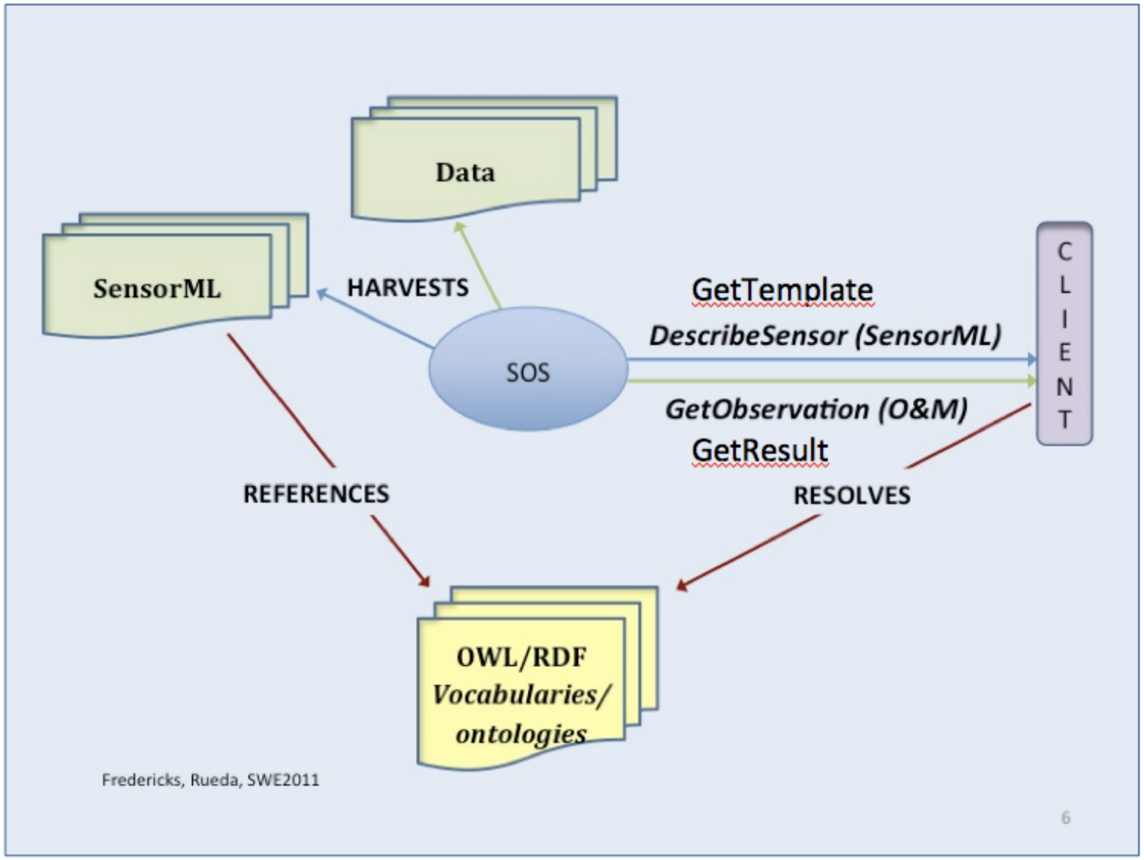
OGC+W3C: SENSOR WEB



Above shows the tasks, tools and interrelationship of the resources.

To facilitate the creation of content and to register the content, the XDOMES team has created:

Ontology Registry and Repository (ORR) where a user can link to existing terms in community adopted ontologies or create their own set of terms to describe their content. (Instructions are on the ESIP Community Page (https://www.esipfed.org/earthcube-xdomes).)



Standards in Sensor Metadata Interoperability

The purpose of linking terms within the SensorML is to enable discovery [3]. For example, if you call your observation "ocean-temperature" and someone is looking for "seawater-temperature", your content will be discoverable.

PROJECT TEAM AND ACKNOWLEDGEMENTS

Janet Fredericks, as the principal investigator (PI) on the Q2O project (http://q2o.whoi.edu), led the development of the quality assessment model [4] that has been used by X-DOMES. She also brings with her many years of research experience, both as an operational oceanographer and a systems programmer. She managed the cabled observatory called the Martha's Vineyard Coastal Observatory, which hosts several research projects each year and has been serving a suite of real-time sensor data since 2001. She served as a liaison to the Inter-Agency Ocean Observation Committee DMAC-ST and also the U.S. IOOS Quality in Real-Time Oceanographic Data Board of Advisors and a participant of the EU-Australia-US Ocean Data Interoperability Platform project. She was involved in EarthCube in its initial phase as a member of the Brokering, Data Access and Semantics & Ontologies community groups and a funded participant in the Layered Architecture Concept Award and participated in the Technical Architecture Committee and the Gap Analysis Working Group of the EarthCube test governance. She has also served two years on the EarthCube Leadership Council.

Mr. Felimon Gayanilo is the Systems Architect for two major scientific data portals in the Gulf of Mexico. The first is the Gulf of Mexico Coastal and Ocean Observing System (GCOOS) that is nested in a National Backbone of coastal observations to aggregate and disseminate the region's near real-time oceanographic data in OGC Sensor Observation Service (SOS) to facilitate interoperability among coastal and ocean observing sensors. The second is the Gulf of Mexico Research Initiative Data and Information Cooperative (GRIIDC), a 10-year multi-disciplinary and multi-institutional research effort in response to the 2010 Deep Horizon oil spill disaster.

Dr. Mike Botts is the author of SensorML and has served as the chair of the OGC® SWE Domain Working Group since its conception. He received the 2008 Gardels Medal for his role in leading the SWE standards activities in OGC®. He is also the lead for development of the current Advanced SensorML Editor and PrettyView, the Space-Time Toolkit visualization package and a variety of open-source libraries in support of SWE. He is currently managing a project to develop an open-source SensorHub (http://opensensorhub.org) to support easy deployment of sensors with immediate access and tasking through SWE 2.0 standards. He also was a Co-PI in the Q2O project. Dr. Botts was an elected member of the OGC Architecture Board (2008-2014).

John Graybeal co-founded the Marine Metadata Interoperability Project in 2004, and continues to serve as the Project Lead. As part of his role he co-hosted several workshops, including the 2006 Sensor Metadata Interoperability workshop. He led MBARI's development of the Shore Side Data System, and guided Data Browser features for Marinexplore/PlanetOS. He served on the W3C's Semantic Sensor Network Working group, served as the co-chair of the ESIP Federation's Attribute Conventions for Data Discovery, and wrote the NetCDF Climate and Forecast Conventions Frequently Asked Questions. He continues to develop and refine vocabularies served by MMI's Ontology Registry and Repository and has sheparded its adoption as a community resource through ESIP (COR) (http://cor.esipfed.org/).

Dr. Carlos Rueda (MBARI) has been the technical lead for the MMI ORR software used for XDOMES, and also used for the MMI Ontology Registry and Repository and ESIP Community Ontology Repository. He coordinated an international group toward the development of a marine device ontology (elements of which were adopted by the W3C Semantic Sensor Network ontology effort) and has been the main developer of the MMI ORR system. He assists IOOS, OOI, ICAN, and other communities in the development of controlled vocabularies and has co-hosted international technical meetings with ontology repository developers, promoting the need to define inter-repository standard interfaces. He has participated in the DataONE Semantics and Integration Working group and the Open Ontology Repository initiative.

Dr. Krzysztof Janowicz is an professor for Geographic Information Science at the Geography Department of the University of California, Santa Barbara, USA as well as one of the two Editors-in-Chief of the Semantic Web journal. He is also the community leader of the 52°North Initiative for Geospatial Open Source Software GmbH semantics community that develops open source solutions that bridge OGC's Geo-Web and the Semantic Web. Janowicz is a member of the W3C SSN-XG that developed the Semantic Sensor Network ontology (SSN), and was responsible for the development of the Stimulus-Sensor-Observation ontology design pattern (SSO) that forms the core of the SSN ontology. Besides ontologies, Janowicz has developed software and specifications for sensor mediation, Semantic Enablement of Spatial Data Infrastructures, and Restful Linked Data proxies for the OGC Sensor Observation Service. He published large Linked Data sets such as the ADL gazetteer and tools for their exploration. He was responsible for the alignment of the X-DOMES SensorML work with the W3C/SOSA efforts.

Acknowledgements

The Q2O project, defining the model for providing information about data quality in a OGC SWE framework, was provided under NOAA’s Cooperative Agreement FY 2007 Regional Integrated Ocean Observing System Development (NOS-CSC-2007-2000875), 2008-2011. The X-DOMES development of tools and registries for the capture and delivery of SensorML and community-adopted vocabularies is funded by the National Science Foundation (NSF) as an EarthCube Integrative Activity. EarthCube is a collaboration between the Division of Advanced Cyberinfrastructure (ACI) and the Geosciences Directorate (GEO) of the US National Science Foundation (NSF). For official NSF EarthCube content, please see: http://www.nsf.gov/geo/earthcube/ (Award #1541008).

WHAT'S NEW AND FUTURE WORK

Whats New?

Integration Of the X-DOMES SRR and with EarthCube GEOCODES

1. The XDOMES home page has the proper GEOCODES tags. If you go to xdomes.org, on the welcome screen or first screen, right-click and view the source, you should see the JSON tags for GeoCODES at the header.

2. The sitemap.xml, now points to all the -generated html data landing pages where Google can come and search (https://xdomes.tamucc.edu/srr/sensorML/sitemap.xml) (https://xdomes.tamucc.edu/srr/sensorML/sitemap.xml)

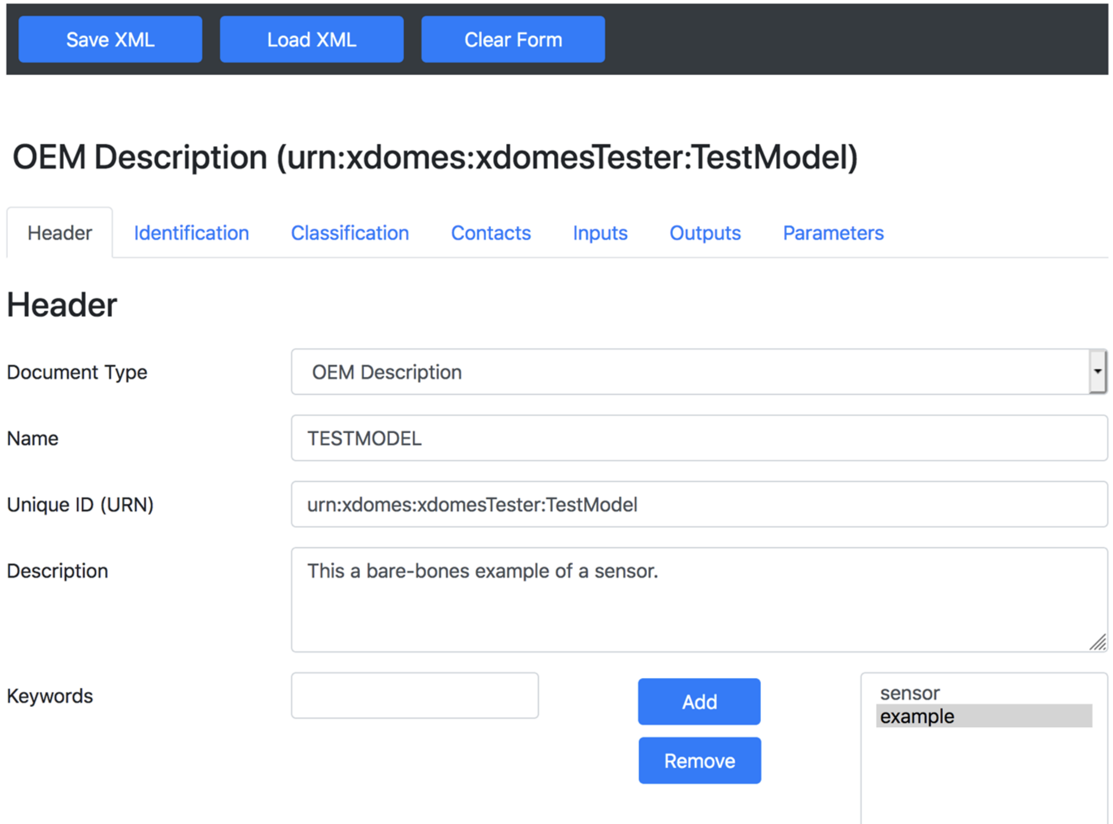
3. Go to the registry (https://xdomes.tamucc.edu/srr/) and click any of the links to any of the data, like the MVCO Workshorse 1200, instead of opening the sensorML XML file directly as what we did originally, it now opens to a landing page (https://xdomes.tamucc.edu/srr/sensorML/urn-whoi-mvco-mvco_workhorse_1200.html). The page provides a hyperlink to the sensorML file for Google crawling.

4. From #3 above, if you view the source (right-click the page and view page source), you will see the inclusion of the required GeoCODES in LD+JSON. These codes are autogenerated as the files are registered.

5. If you go to Google's DataSearch engine (https://toolbox.google.com/datasetsearch) and test, like entering "gill windsonic", it will now give you a link to xdomes' registry.

BASIC EDITOR

The Basic Editor was added to our toolset to provide an easier way to get started.



Future Work

Earth Science communities need to continue to work towards developing terms and ontologies (ORR), including observable properties, sensor parameters, processing descriptions, sensor capabilities and characteristics and observeration parameters. These terms should be resolved across-domains for interdisciplinary discovery and access.

Manufacturers need to be encouraged to participate - taking their technical spec sheets from human-readable forms to machine-actionable, persistent documents.

The Basic Editor needs to be updated with more category headings, (See italicized items in the SensorML Contents list in the panel to the left.)

The more advanced editor needs to find a stable home. It keeps disappearing on us as its broader funding changes.

NOTE: Persistence is a major issue for interdisciplinary resources. Perhaps ESIP is a place to have information that enables interoperable access to data. It is the Earth Science Information Partnership - so perhaps it could broaden its information from communication to interdisciplinary tools such as registries and repositories. They needn't host data but could provide the bridges across domains, disciplines, and governing bodies to enable research into the Earth as a system.

CV

Janet Fredericks is an Emeritus Research Scholar with the Woods Hole Oceanographic Institution (WHOI). She has spent more than 30 years at WHOI in geo-chemistry, geophysics and coastal ocean fluid dynamics as an information systems programmer and operational oceanographer. https://www.whoi.edu/profile/jfredericks/ (https://www.whoi.edu/profile/jfredericks/)

Felimon Gayanilo is a Systems Architect/Enterprise IT working on various projects with the Harte Research Institute. He has over three decades of experience in the design, development, and deployment of information systems in a local, national, regional and international settings before he joined Texas A&M University in 2012. https://www.harteresearchinstitute.org/people/felimon-gayanilo (https://www.harteresearchinstitute.org/people/felimon-gayanilo)

ABSTRACT

With observational data becoming widely available, researchers struggle to find information enabling assessment for its reliable use. A small first-step toward enabling data quality assessment of observational data is to associate the data with the sensor used to make the observations and to have the sensor description machine-harvestable. In the latest additions to the X-DOMES (Cross-Domain Observational Metadata for Enviromental Sensing) toolset, we have created targeted editors for creating SensorML documents to describe sensor models. The team has adjusted its delivery to enable integration of the X-DOMES content with the GEOCODES (JSON-LD/schema.org) EarthCube project. At our poster-session, we will highlight the new changes and capabilities and demonstrate the use of new X-DOMES tools.

REFERENCES

[1] Fredericks, J. (2015). Persistence of knowledge across layered architectures. In P. Diviacco, P. Fox, C. Pshenichny, and A. Leadbetter (Editors), Collaborative Knowledge in Scientific Research Networks. Hershey, PA: IGI Global, 20 pp http://www.igi-global.com/book/collaborative-knowledge-scientific-research-networks/110007

[2] Fredericks, J., Botts, M. Promoting the capture of sensor data provenance: a role-based approach to enable data quality assessment, sensor management and interoperability. Open geospatial data, softw. stand. 3, 3 (2018). https://doi.org/10.1186/s40965-018-0048-5

[3] Graybeal, J. & Coauthors (2012). "Semantic mediation of vocabularies for ocean observing systems", Computers & Geosciences, Vol. 40, March 2012, Pages 120–131 http://www.sciencedirect.com/science/article/pii/S0098300411002603

[4] 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, (2009). "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 WPP-306, doi:10.5270/OceanObs09.cwp.31