Earth Science Information Partners (ESIP) EnviroSensing Cluster

Don Henshaw[1], Janet Fredericks[2], and Alison Adams[3]

[1] U.S. Forest Service Pacific Northwest Research Station, [2] Woods Hole Oceanographic Institution, [3] University of Vermont

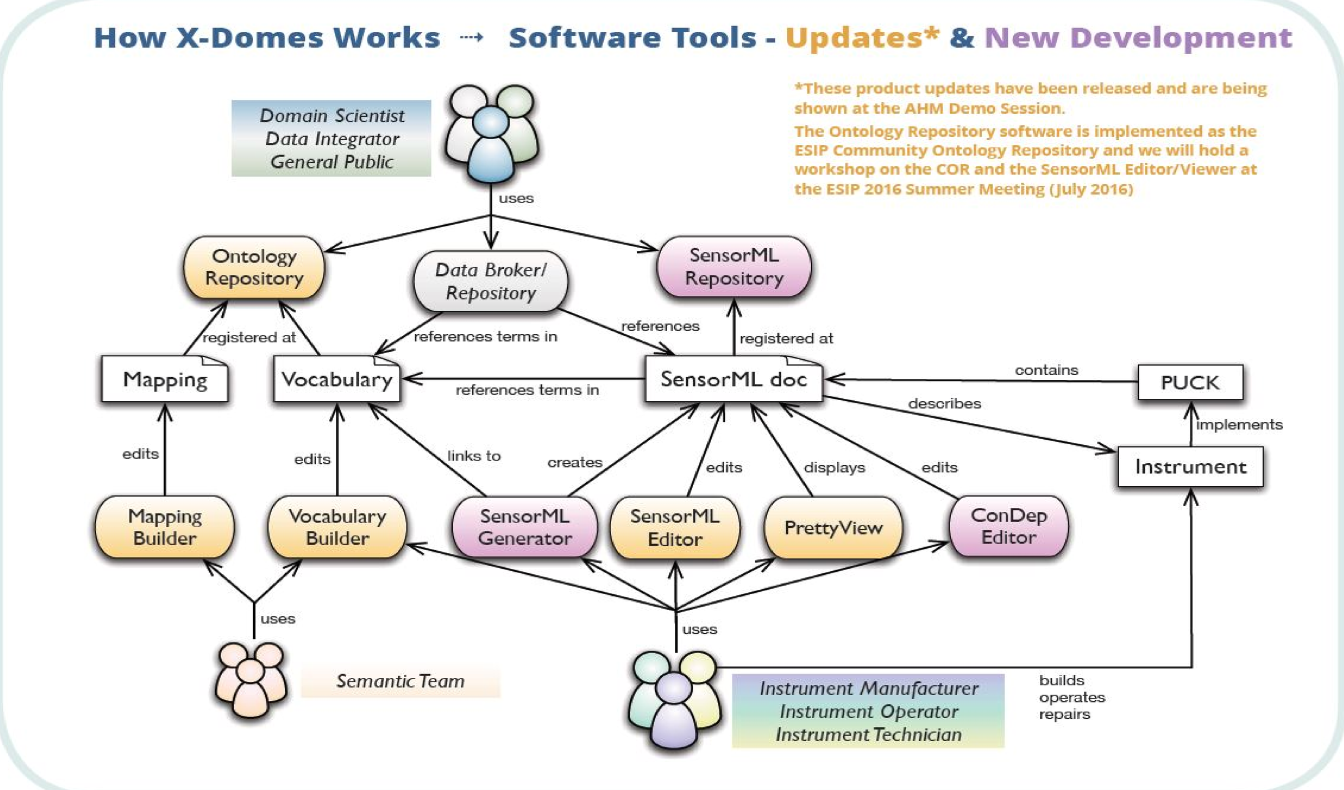
EnviroSensing Cluster

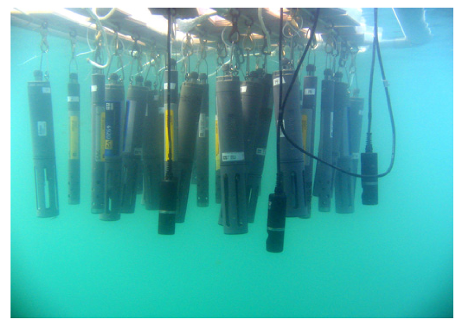
Current emphasis

X-DOMES project (Cross-Domain Observational Metadata Environmental Sensing) funded by NSF as part of the EarthCube Integrative Activities

X-DOMES primary goal

- Work with sensor manufacturers to encourage them to begin to fully-describe sensors using a combination of controlled-vocabularies, OGC-SWE, SensorML and the Semantic Web





Capture Metadata using Semantics Web (W3C) and SensorML (OGC-SWE)

- Domain Communities create registered vocabularies and ontologies in an Ontology Registry for versioning
- Sensor manufacturers create the Original Equipment Manufacturer (OEM) SensorML documents that describe sensor model metadata
- Field operators create SensorML documents that reference the OEM SensorML and provide information about the specific sensor and its deployment.
- Data providers can reference these files and include SensorML that describes processes, such as QC and derived products.

Access and Integrate

- Integration within BCO-DMO, R2R, DataONE, CUAHSI, CHORDS, GeoLINK and LTER infrastructure will help us document functionality, feasibility and relevance to goals
- Document the ability to translate and transfer content through brokers to demonstrate interoperability
- The SensorML editor enables the inclusion of links to the registered terms
- Content Access Modules (Python/Matlab) enable access to the registered SensorML documents and the associated RDF/OWL links.

Community Engagement for Sustainability

- Define requirements for sustainability
- Working within the EarthCube Community
- Engaging the Earth Science Information Partnership (ESIP) EnviroSensing Cluster
- Planning with EarthCube and other existing data facilities

X-DOMES Team: Janet Fredericks (WHOI), Mike Botts (Bott, Inc.), Felimon Gayanilo (Texas A&M), John Graybeal (MMI), Krzysztof Janowicz (UCSB), Carlos Rueda (MBARI)

Sensor and Sensor Data Management Best Practices

http://wiki.esipfed.org/index.php/EnviroSensing_Cluster

Sensor, site, and platform selection

- Selection of sites, science platforms and support systems are interacting planning processes
  - Communication among PI's, techs, and information managers
- Data quality and longevity is ultimate goal
  - Robust and widely-used core systems and sensors
  - Standardize sensor and support hardware, software, designs
- Optimal siting for science objectives can be impeded
  - land ownership/permitting, seasonal weather patterns, logistical access, availability of services (e.g., power sources, communications), operating budget

Data acquisition and transmission

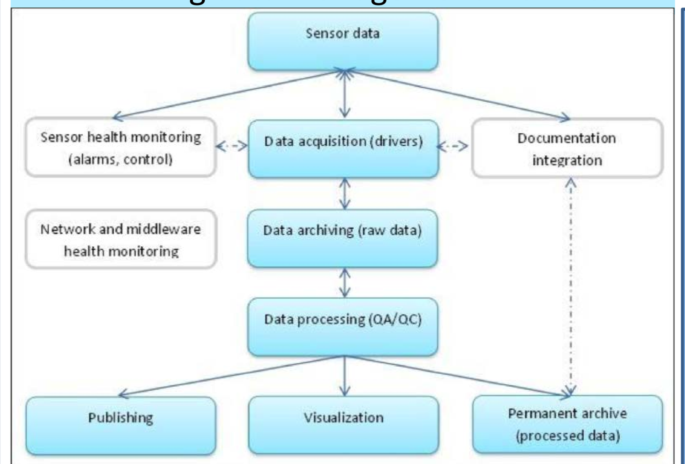
- Manual downloads of sensor data
  - May not be sufficient to assure data security
  - Does not allow direct control of devices
- Remote data acquisition considerations:
  - Collection frequency and need for immediate access
  - Uni- versus bi-directional transmission methods
  - Bandwidth requirements to transfer the data
  - Line-of-site communication or repeaters
  - Hardware and network protocols
  - Power consumption of the system components
  - Physical and network security requirements
  - Reliability and redundancy
  - Expertise
  - Budget

Sensor management, tracking, and documentation

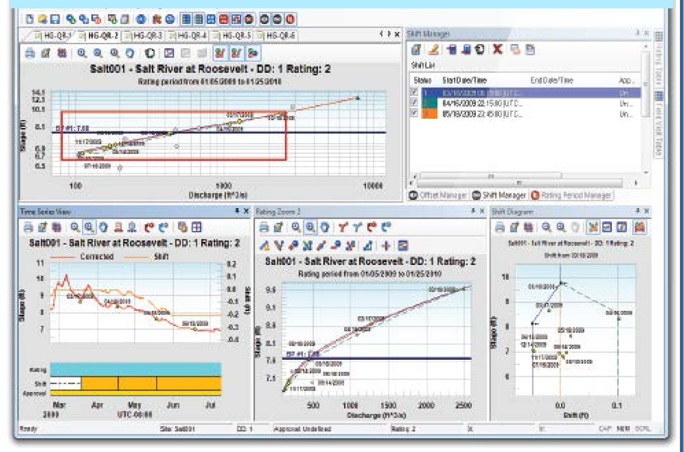
- Documentation of field procedures and protocols:
  - Site visits, sensor tracking, calibration and maintenance activities, datalogger programs
- Sensor event tracking
  - Sensor event histories are essential for internal review of data, e.g., sensor failures, disturbances, method changes
- Integration of sensor documentation with the data
  - Associate data qualifier flag with each data value
  - Add a "methods_code" data column for easy user identification of methodology changes for a given sensor
- Communication between field and data personnel
  - Example field note database:

SiteID	Datalogger ID	SensorID	date time begin	date time end	category	notes	Note taker
					controlled vocabulary		

Streaming data management workflow

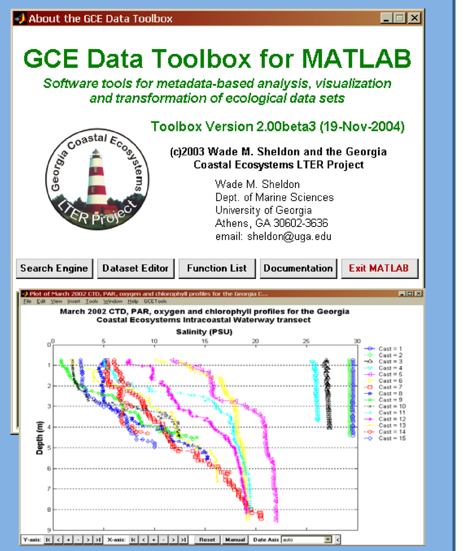


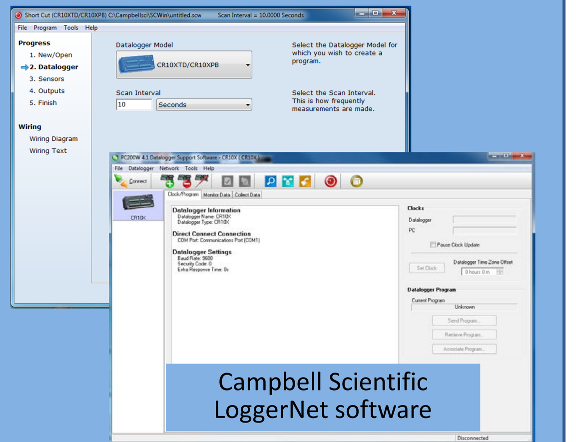
Aquarius software



Streaming data management middleware

- "Middleware" software packages and procedures
  - Enable communication and management of data between field sensors and a client such as a database, website or software application
  - Purposes include the collection, archival, analysis, and visualization of data
  - Middleware is often chained together into a scientific workflow to meet multiple functional requirements
  - Considerations:
    - Licensing, cost, interoperability of components
- Proprietary middleware / software
  - Campbell Scientific – LoggerNet
  - Aquatic Informatics – Aquarius
  - Vista Engineering – Vista Data Vision (VDV)
  - YSI – EcoNet
  - NexSens Technology – WQData Live
- Open source environments for middleware
  - GCE Data Toolbox (MATLAB required)
  - CUAHSI Hydrologic Information System (HIS)
  - DataTurbine Initiative





Campbell Scientific LoggerNet software

Sensor data quality assurance and quality control (QA/QC)

- Quality assurance – preventative measures
  - Routine calibration and maintenance
    - Anticipate common repairs and replacement parts
  - Design
    - Assure proper installation and protection
    - Sensor redundancy
  - Regular human inspection and evaluation of sensor network
    - Automated alerts; in situ webcams
- Quality control – checks in near real-time
  - Timestamp integrity (Date/time)
  - Range checks
  - Internal (plausibility) checks
  - Variance checks / Outlier detection
  - Persistence checks
  - Spatial checks / Correlations with nearby sensors

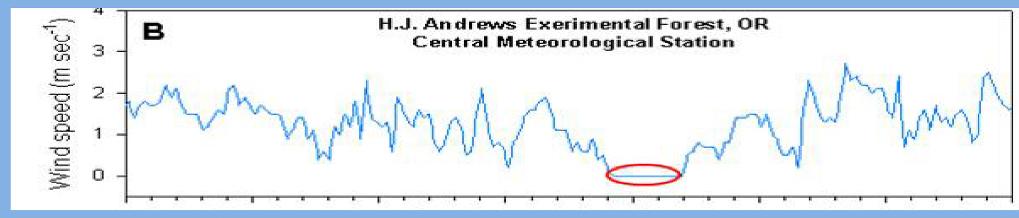


Image from Campbell et.al., Bioscience, 2013.

Sensor data archiving

- Archiving strategies
  - Create well documented data snapshots
  - Assign unique, persistent identifiers
  - Maintain data and metadata versioning
  - Store data in text-based formats
- Partner with cross-institution supported archives
  - Federated archive initiatives such as DataONE
  - Community supported, e.g., the LTER NIS
- Best practices
  - Develop an archival data management plan
  - Implement a sound data backup plan
  - Archive raw data (but they do not need to be online)
  - Make data publicly available
    - Assure appropriate QA/QC procedures are applied
  - Assign QC level to published data sets