SensorDat: a testbed for improving real-time sensor data quality and provenance

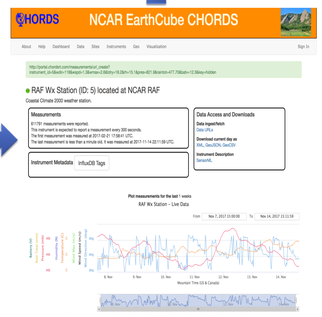
Mike Daniels (NCAR), Matt Bartos (UMich), Connor Scully-Allison (UNR), Scotty Strachan (UNR), Renée Brown (UNM), Aaron Botnick (NCAR), Charlie Martin (NCAR)

SensorDat is an ESIP Lab Incubator project which was funded in the summer of 2018. Real-time sensors are increasingly being used for scientific analysis and discovery in earth science research. The Internet of Things (IoT) concept has led to an environment in which small, inexpensive sensors are becoming ubiquitous, providing a whole new set of real-time measurements to the research community. Sensor data used for scientific research purposes require additional sophistication due to issues such as interoperability and metadata requirements, spatial and temporal coverage, data quality considerations and precise measurement specifications. Through systems like CHORDS, it is now possible to bring a level of standardization and consistency to these new sensor streams, addressing data quality issues in real-time so that problems are caught quickly, ultimately improving these measurements. In addition, CHORDS adheres to evolving metadata standards and controlled vocabularies to help researchers discover streaming data in their areas of interest while fully describing the measurements being taken (e.g., variables measured, units of measurement, spatial and temporal coverage, etc.). Through the ESIP lab, we have 1) extended the use of CHORDS to real-time data streams that are outside of the traditional NSF Geosciences domain, including new sensors that take advantage of IoT miniaturization and 2) develop advanced workflow prototypes focused on automated data quality recipes and annotation.

CHORDS creates real-time streams that are "Born Connected"[1]

SensorML Instrument description

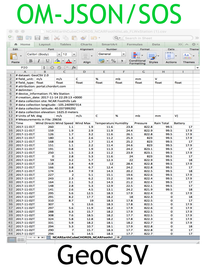
Sensor Data



CHORDS instance

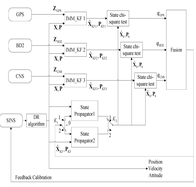
GeoJSON

JSON-LD and OM-JSON/SOS



GeoCSV

Archives



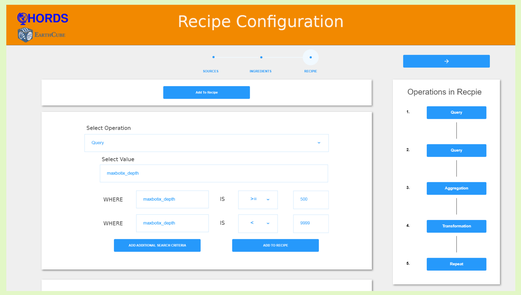
Algorithms

Prediction Models

CHORDS uses standards and controlled vocabularies where they exist. Items in green are future developments.

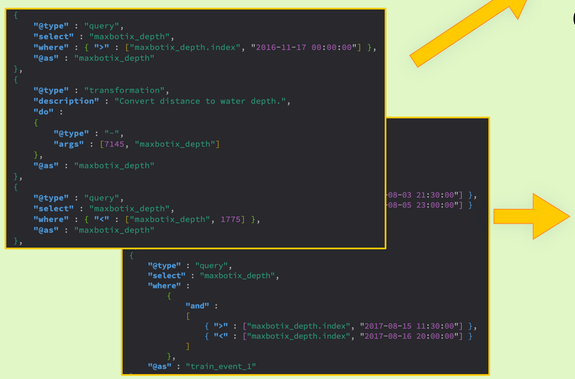
[1]Leadbetter, Adam & Smyth, Damian & Fuller, Robert & Shepherd, Adam & O'Grady, Eoin. (2016). Where Big Data meets Linked Data: Applying standard data models to environmental data streams. 10.1109/BigData.2016.7840943.

Data quality "recipes" applied to data streams

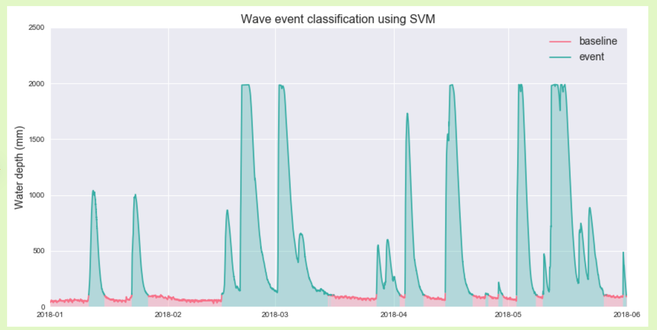


Quality Control recipe selector

Raw data (top) and corrected data (bottom) after the Quality Control recipe has been applied



JSON-LD description of Quality Control recipe



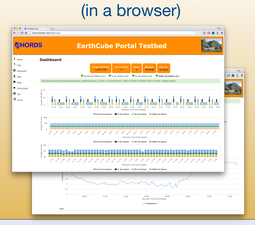
JSON-LD documents specify training and test data for machine learning classification

SensorDat's Real-time QC Workflow Prototype

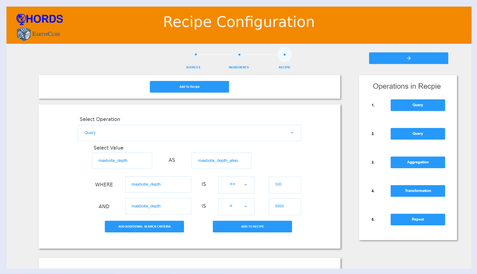
Adapt your instrument to send data via simple HTTP queries (it's not that hard)

http://myportal.org/data?speed=21&dir=355

Monitor your system and instruments from anywhere (in a browser)



{"wdir":[75.0],"wspd":[2.5],"wmax":[6.4],"tdry":[26.3]}



CHORDS is funded by the US National Science Foundation EarthCube program, grants #1639750, 1639720, 1639640, 1639570 and 1639554. This work is based on materials provided by the ESIP Lab with support from the National Aeronautics and Space Administration (NASA), National Oceanographic and Atmopsheric Administration (NOAA) and the United States Geologic Survey (USGS).

More information about CHORDS: See http://chordsrt.com

For more information about SensorDat: See http://github.com/ESIPFed/SensorDat

See more live data: http://portal.chordsrt.com