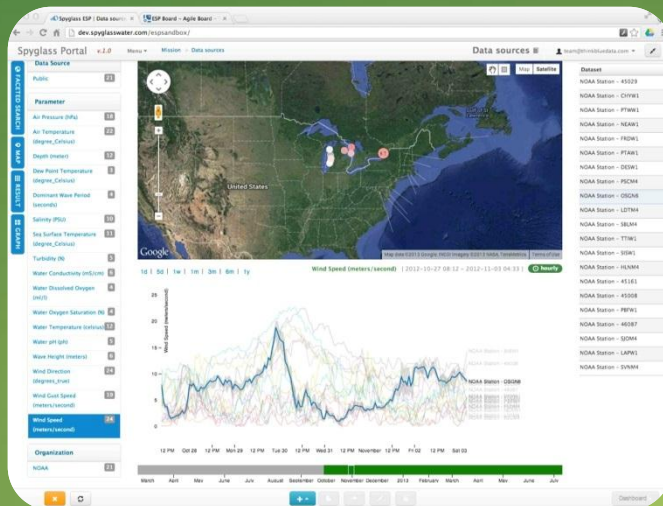




# AIT Technology Event



## Seeing the Big Picture

*Sensing, Linking, Analyzing and Visualizing Big Data*

*Dr. Paul Janecek*





# Content

- Introduction
  - Seeing Details and Context in Big Data
- Example: Real-Time Monitoring
  - Sensing: In-Situ Data
  - Linking: IOOS
  - Analyzing: Pattern Recognition & Analysis
  - Visualizing: Data Portal
- Challenges



# BIG DATA

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

The "three V's", i.e the Volume, Variety and Velocity of the data coming in is what creates the challenge.

## VOLUME



## VARIETY

**PEOPLE TO PEOPLE**  
NETIZENS, VIRTUAL COMMUNITIES, SOCIAL NETWORKS, WEB LOGS...

**PEOPLE TO MACHINE**  
ARCHIVES, MEDICAL DEVICES, DIGITAL TV, E-COMMERCE, SMART CARDS, BANK CARDS, COMPUTERS, MOBILES...

**MACHINE TO MACHINE**  
SENSORS, GPS DEVICES, BAR CODE SCANNERS, SURVEILLANCE CAMERAS, SCIENTIFIC RESEARCH...

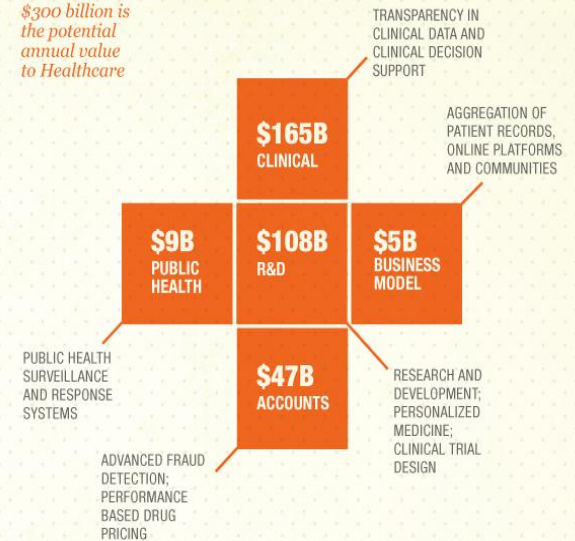
## VELOCITY

**2.9 MILLION**  
EMAILS SENT EVERY SECOND

**20 HOURS**  
OF VIDEO UPLOADED EVERY MIN

**50 MILLION**  
TWEETS PER DAY

\$300 billion is the potential annual value to Healthcare

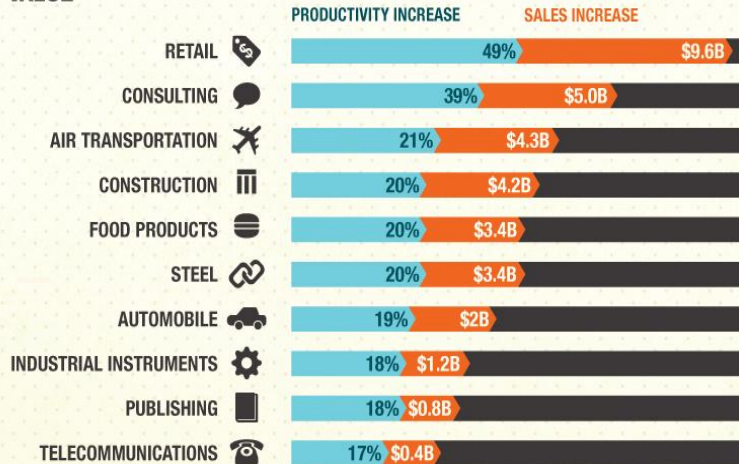


## VALUE

57.6% OF ORGANIZATIONS SURVEYED SAY THAT BIG DATA IS A CHALLENGE

72.7% CONSIDER DRIVING OPERATIONAL EFFICIENCIES TO BE THE BIGGEST BENEFIT OF A BIG DATA STRATEGY

50% SAY THAT BIG DATA HELPS IN BETTER MEETING CONSUMER DEMAND AND FACILITATING GROWTH



40% PROJECTED GROWTH IN GLOBAL DATA CREATED PER YEAR



5% PROJECTED GROWTH IN GLOBAL IT SPENDING PER YEAR

The estimated size of the digital universe in 2011 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44 fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

Sources: 1 Reaping the Rewards of Big Data - Wipro Report 2 Big Data: The Next Frontier for Innovation, Competition and Productivity - McKinsey Global Institute Report 3 CamScore, Radical Group 4 Measuring the Business Impacts of Effective Data - study by University of Texas, Austin 5 US Department of Labour

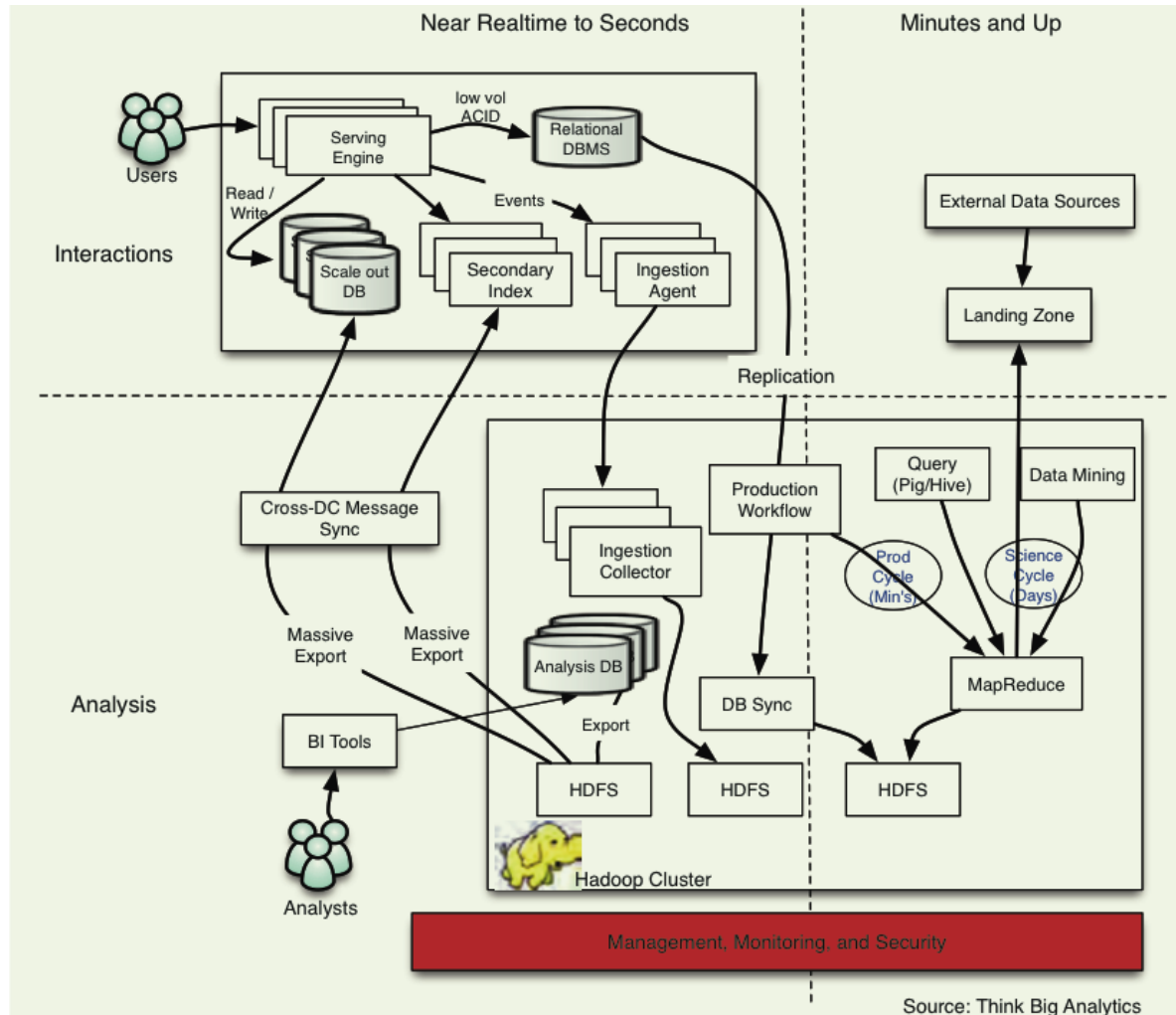
DO BUSINESS BETTER

NYSE:WIT | OVER 130,000 EMPLOYEES | 54 COUNTRIES | CONSULTING | SYSTEM INTEGRATION | OUTSOURCING





# Example Big Data Architecture



Source: Think Big Analytics





# Example: In-Situ Real-Time Monitoring

- **Detecting Hazardous Algal Blooms (HAB)**
- **In-Situ Sensors**
  - Environmental Sample Processor (ESP)
- **Networked Ocean Data**
  - Integrated Ocean Observing Initiative (IOOS)
  - Ocean Observatories Initiative (OOI)
- **Data Visualization Portal**
  - Spyglass Data Portal
    - Developed by Think Blue Data, Thailand
- **Impact:**
  - Detection reduced from 3-5 **days** to 3.5 **hours**





# Big Data Challenges

- **Variety**
  - Structured and Unstructured data
    - Time series, Log Files, Images
- **Volume**
  - Massive historical archives of open data
- **Velocity**
  - Real-time and near-real-time data streams



# Sensing: In-Situ

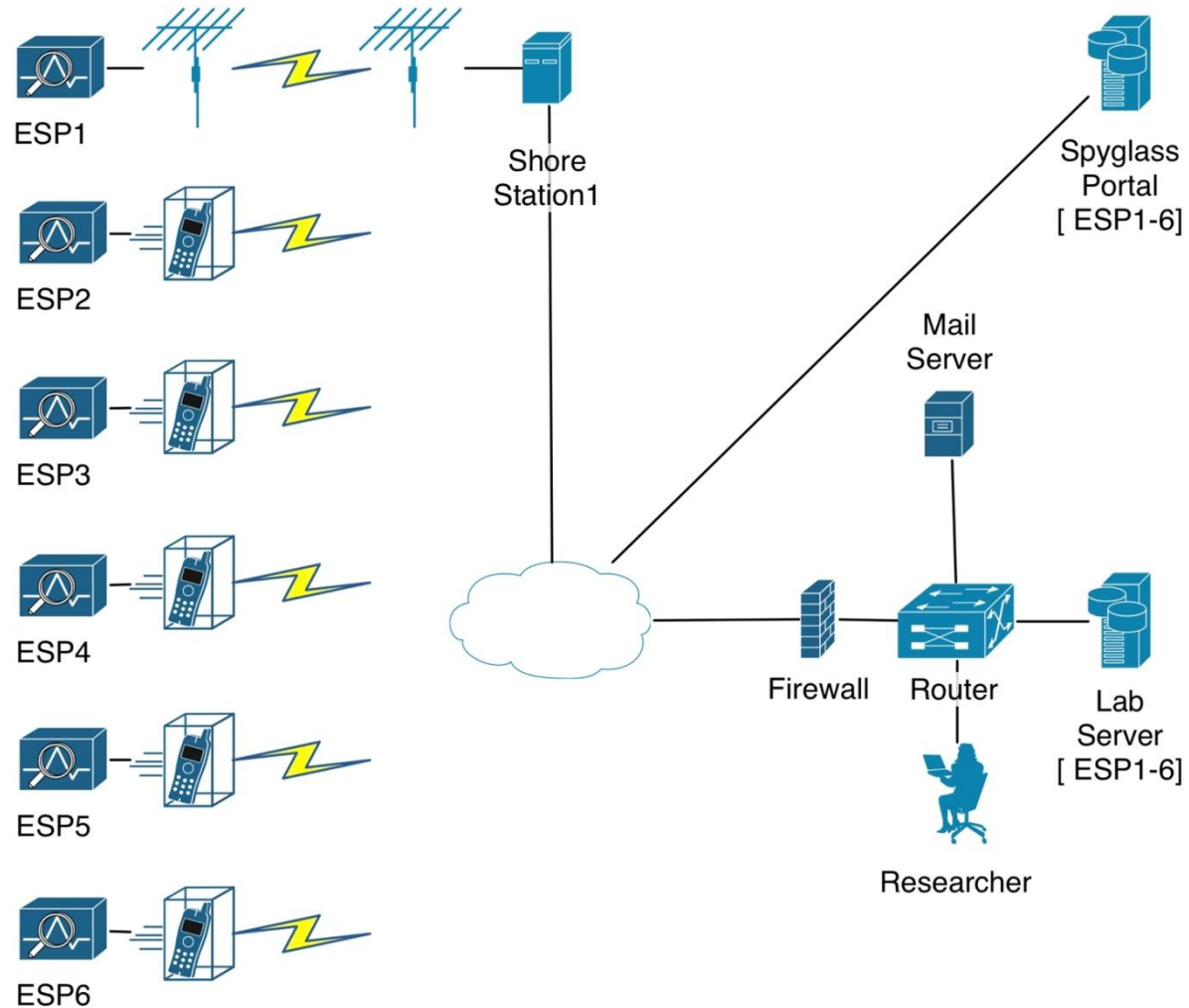
- **ESP: an Underwater Ecogenomic Robot**
  - Detects DNA and Toxins
  - Sends result as image
  - Near real-time data



# Sensing:

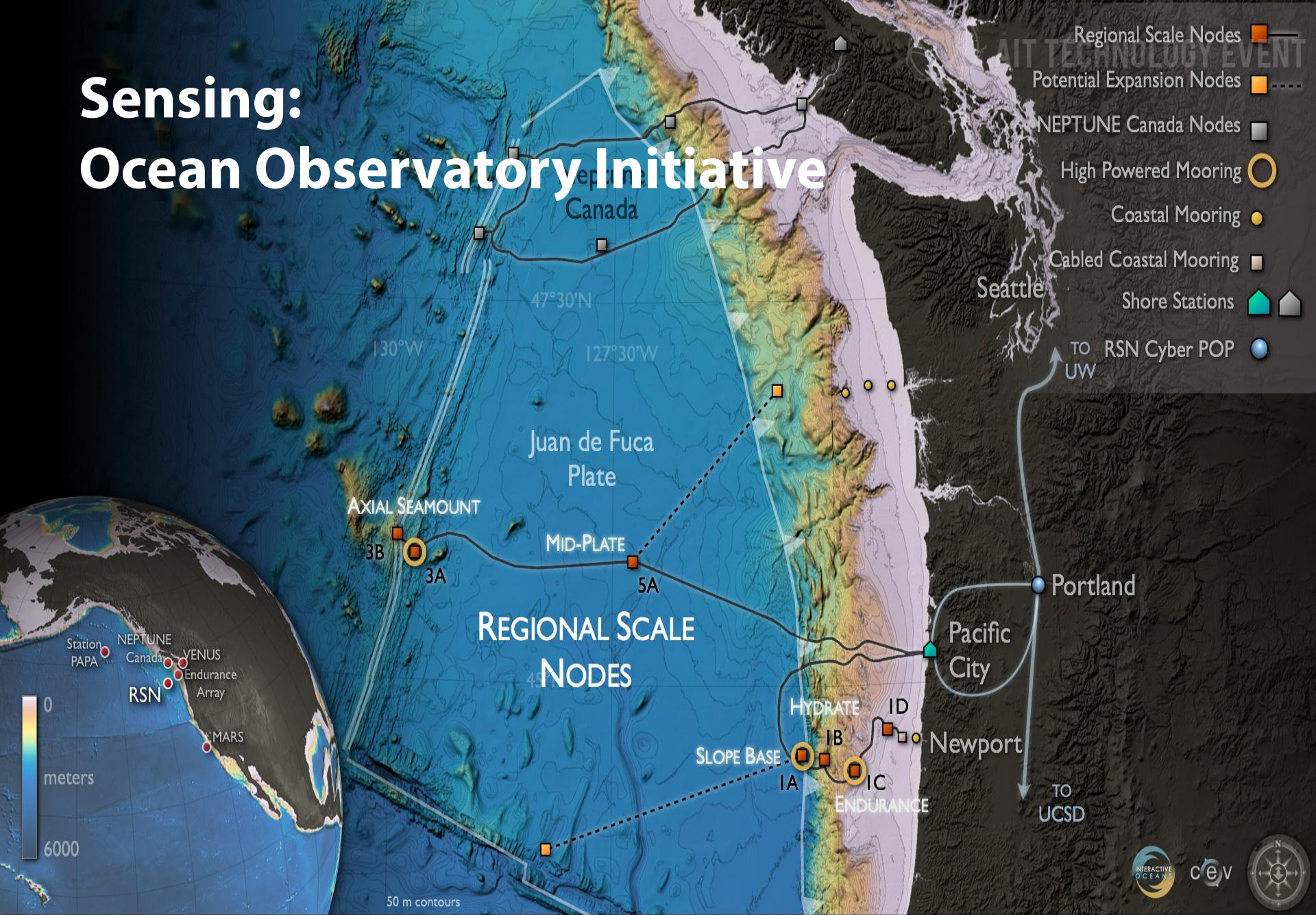


# In-Situ Sensor Network





# Sensing: Ocean Observatory Initiative

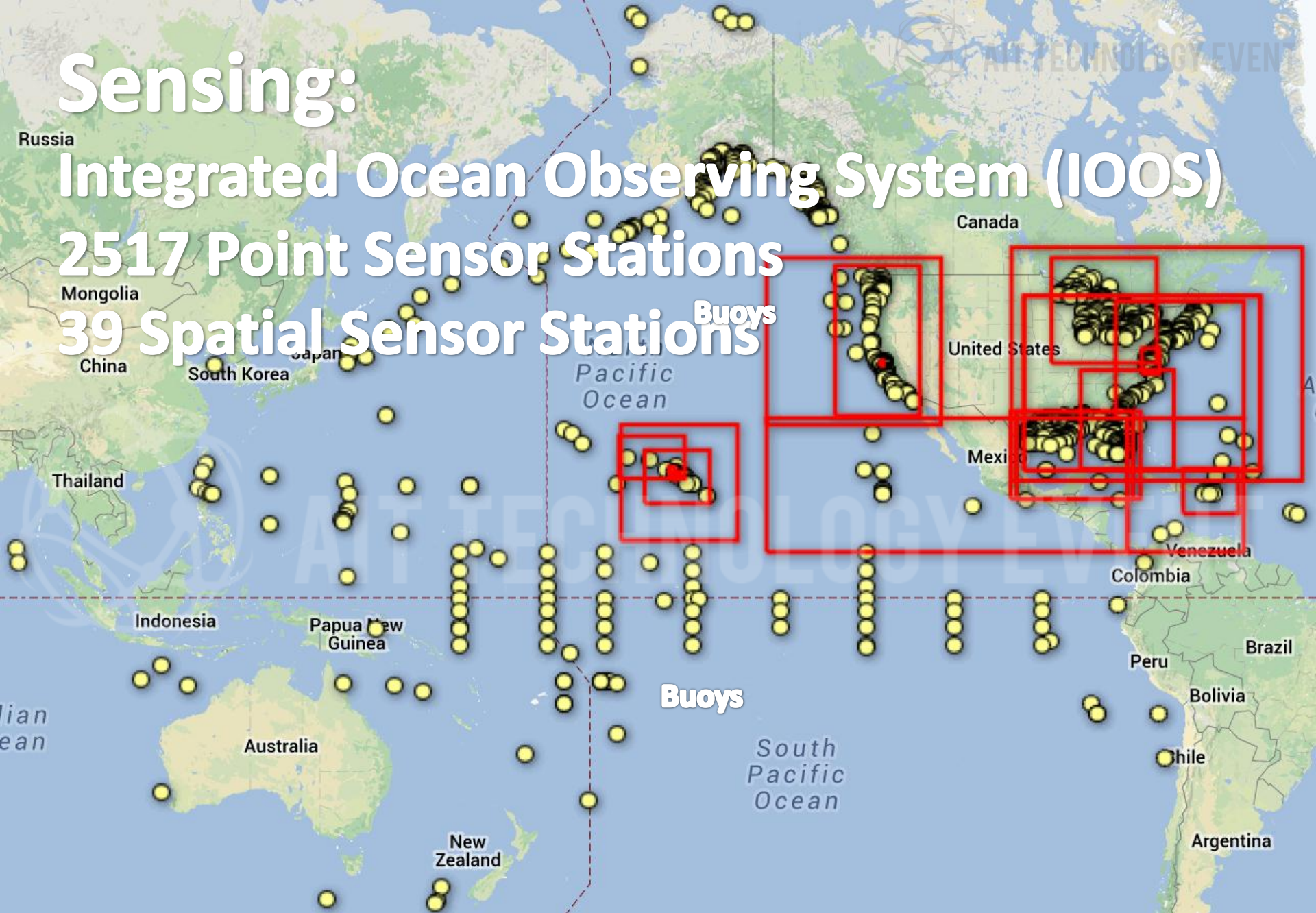


# Sensing:

## Integrated Ocean Observing System (IOOS)

2517 Point Sensor Stations

39 Spatial Sensor Stations



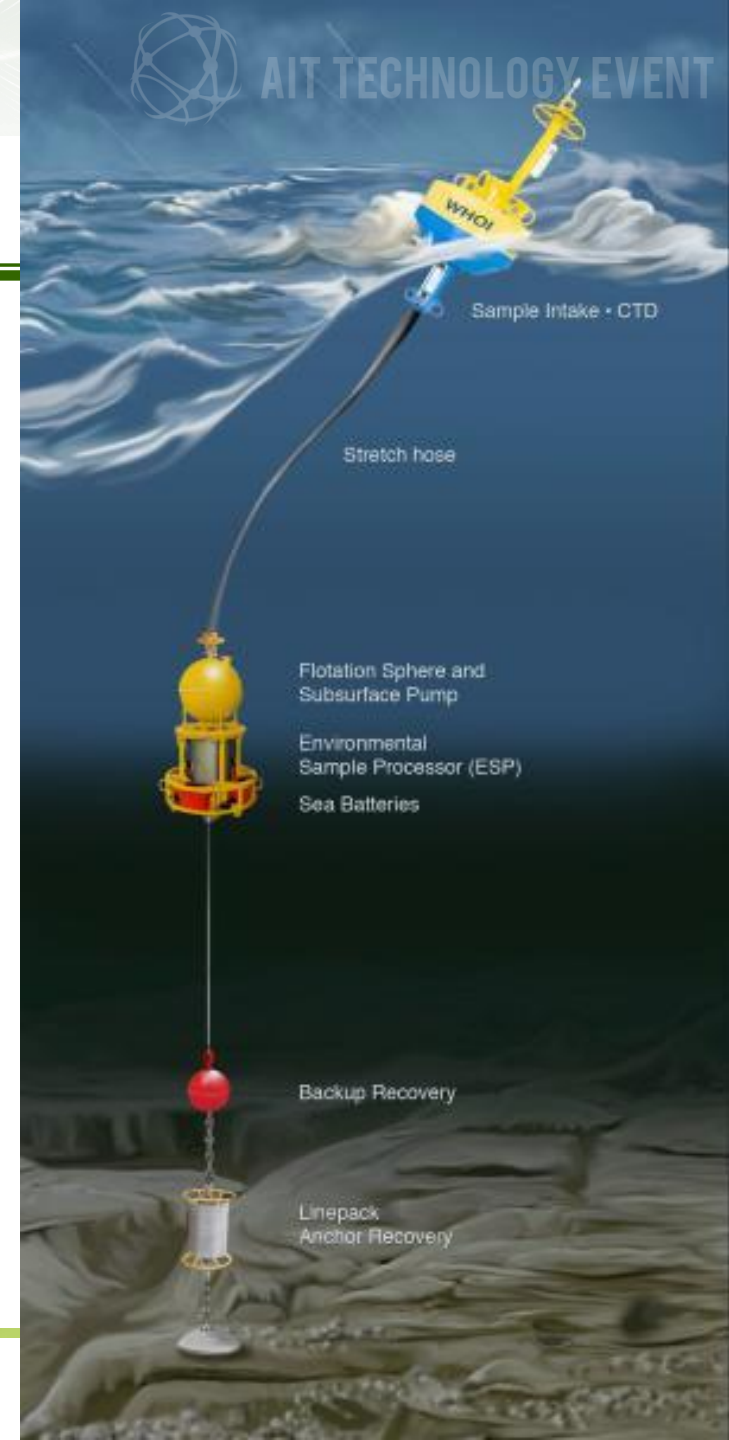
# Sensing: Challenges

- **Data Sources**

- Data Capture
- Data Storage Standards
- Metadata
- Quality

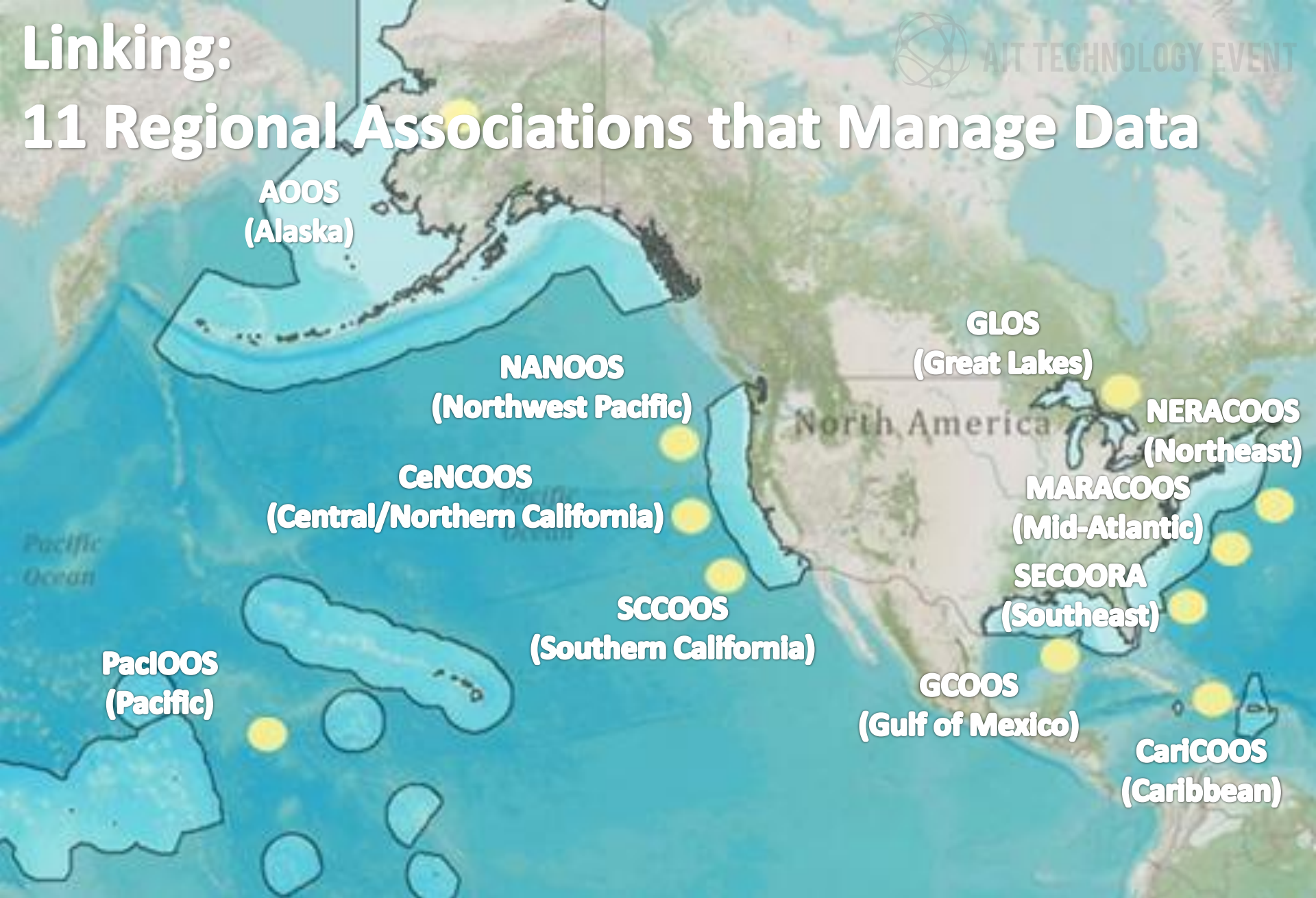
- **Data Access**

- Networking
- Security



Linking:

# 11 Regional Associations that Manage Data





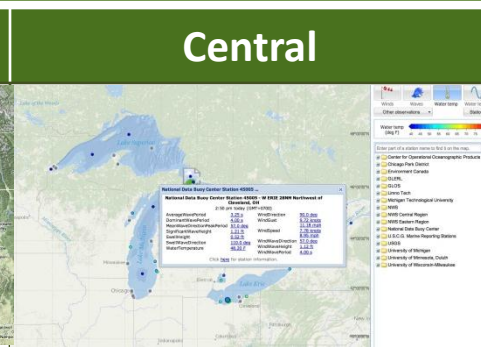
# Linking: 11 IOOS Regional Associations



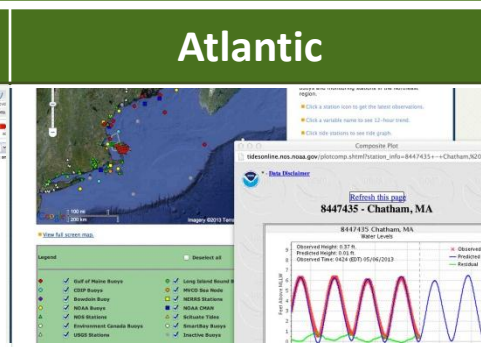
Pacific



Pacific Coast



Central



Atlantic

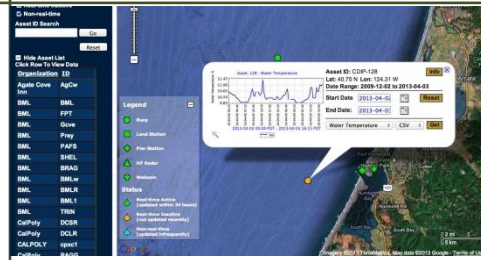
AOOS (Alaska)

NANOOS (Northwest Pacific)

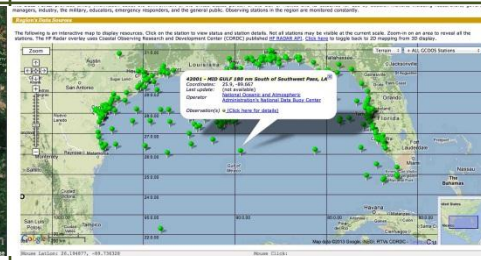
GLOS (Great Lakes)

NERACOO (Northeast)

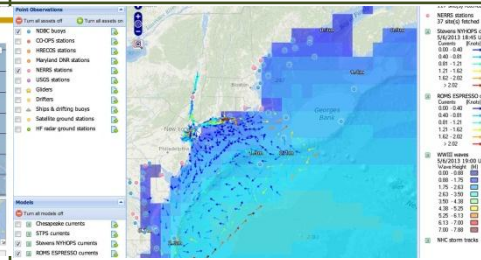
**Integrate,  
Catalog,  
Provide Public Access**



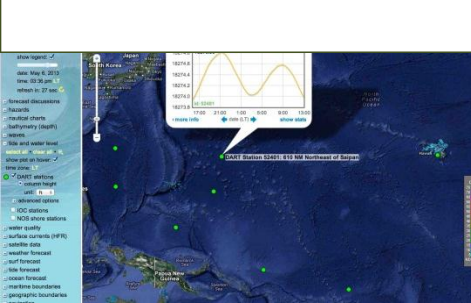
CeNCOOS (Central California)



GCOOS (Gulf of Mexico)



MARACOO (Mid-Atlantic)



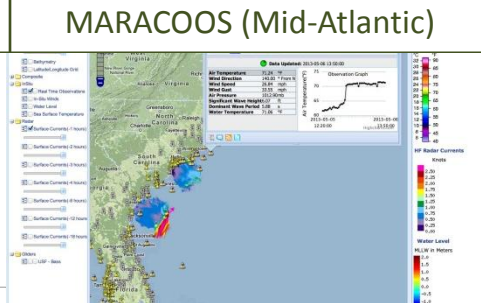
PaCLOOS (Pacific)



Southern California



CariCOOS (Caribbean)

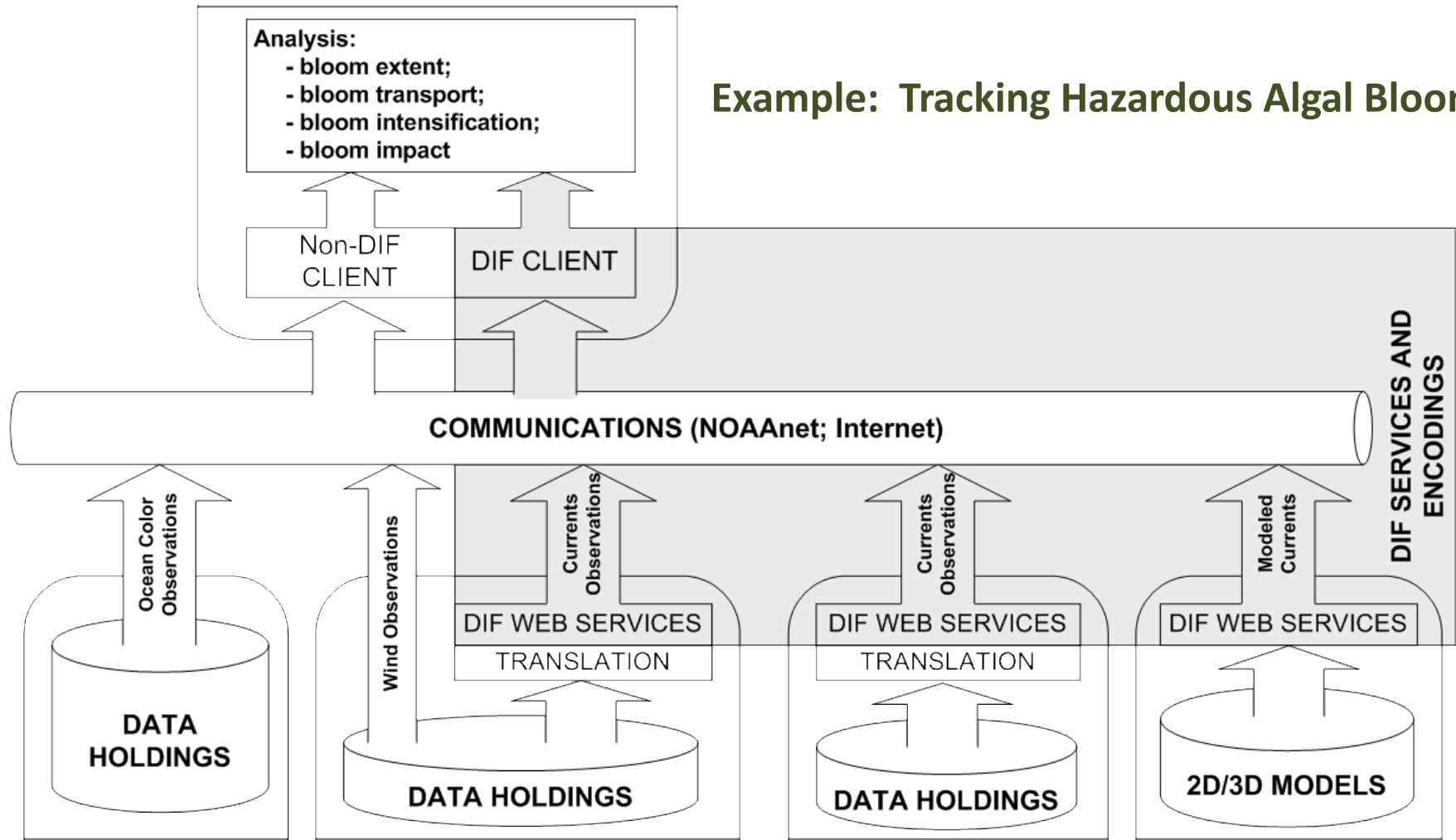


SECOORA (Southeast)



# Linking: IOOS Data Integration Framework

## Example: Tracking Hazardous Algal Blooms

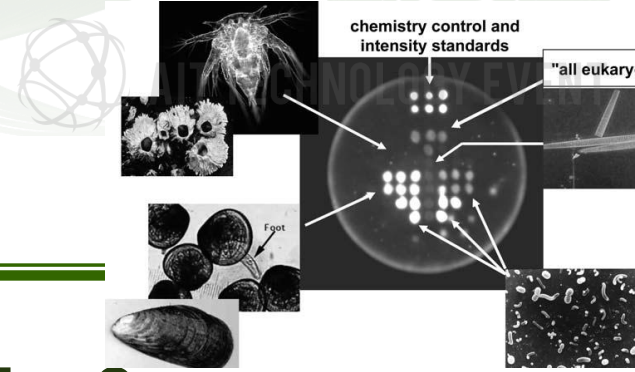


# Linking: Challenges

- **Data Access**
  - Data Discovery
  - Standards
- **Data Integration**
  - Vocabularies
  - Data Preparation

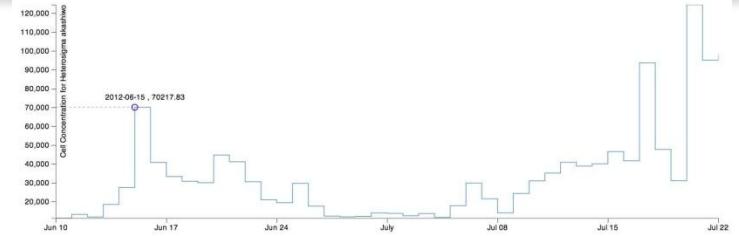
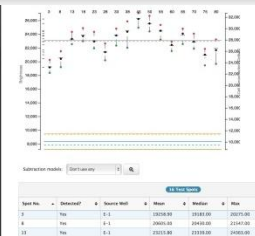
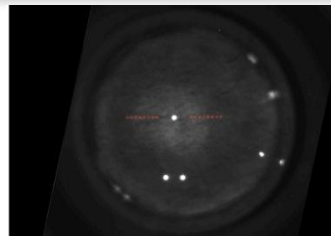
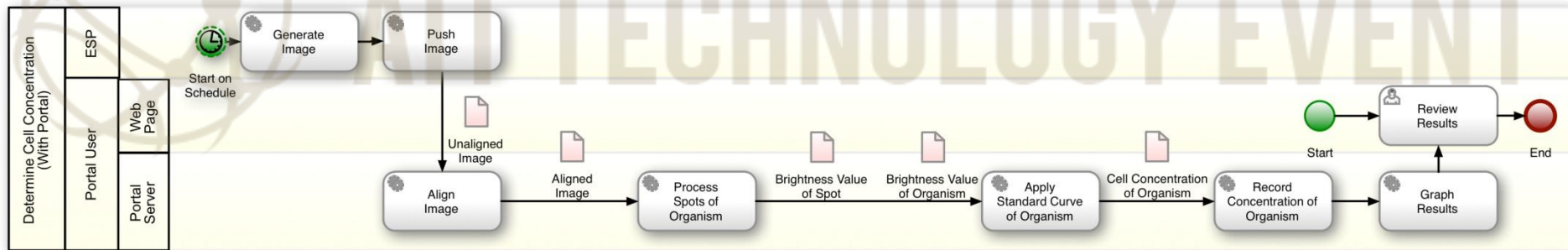


# Analyzing: In-Situ



- **Actionable Data from In-Situ Sensors**

- Parsing, Data Preparation
- Pattern Recognition, Signal Detection
- Update Monitoring Dashboard

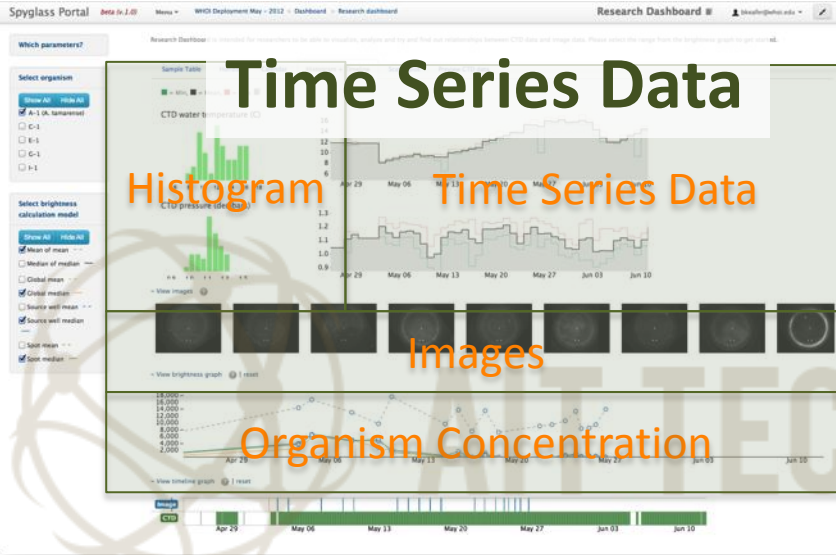






# Visualizing: Single Source Data

## ESP Sensor Data



## Time Series Data

Histogram

Time Series Data

Images

Organism Concentration

## Image Data

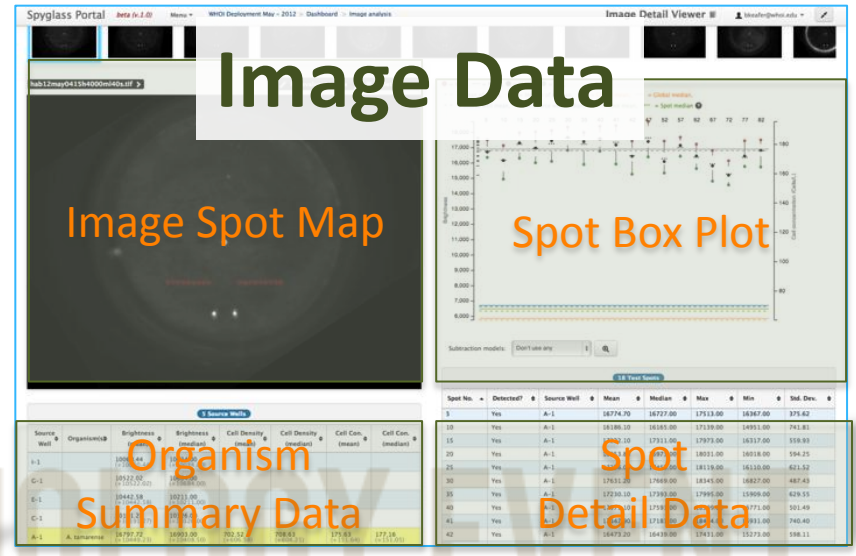


Image Spot Map

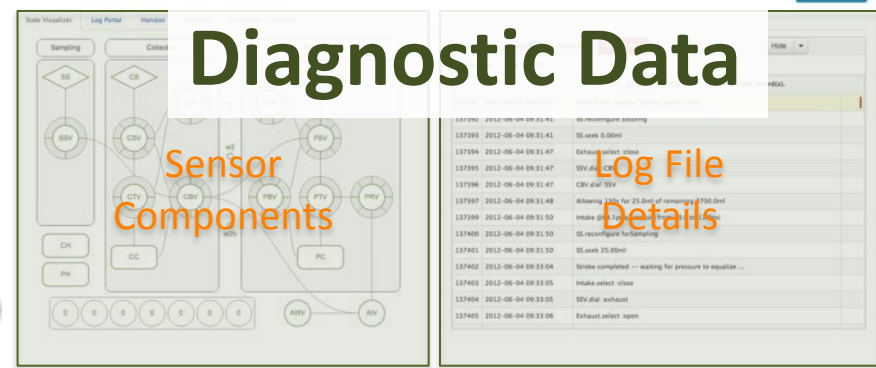
Organism

Summary Data

Spot

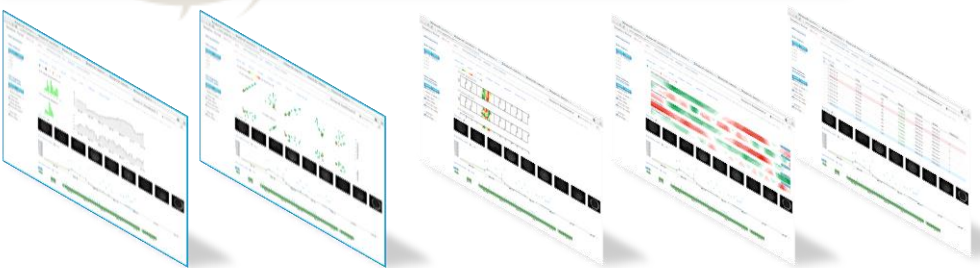
Detail Data

## Diagnostic Data



Sensor Components

Log File Details



Time Series    Scatterplot Matrix    Calendars    Horizon Chart    Details

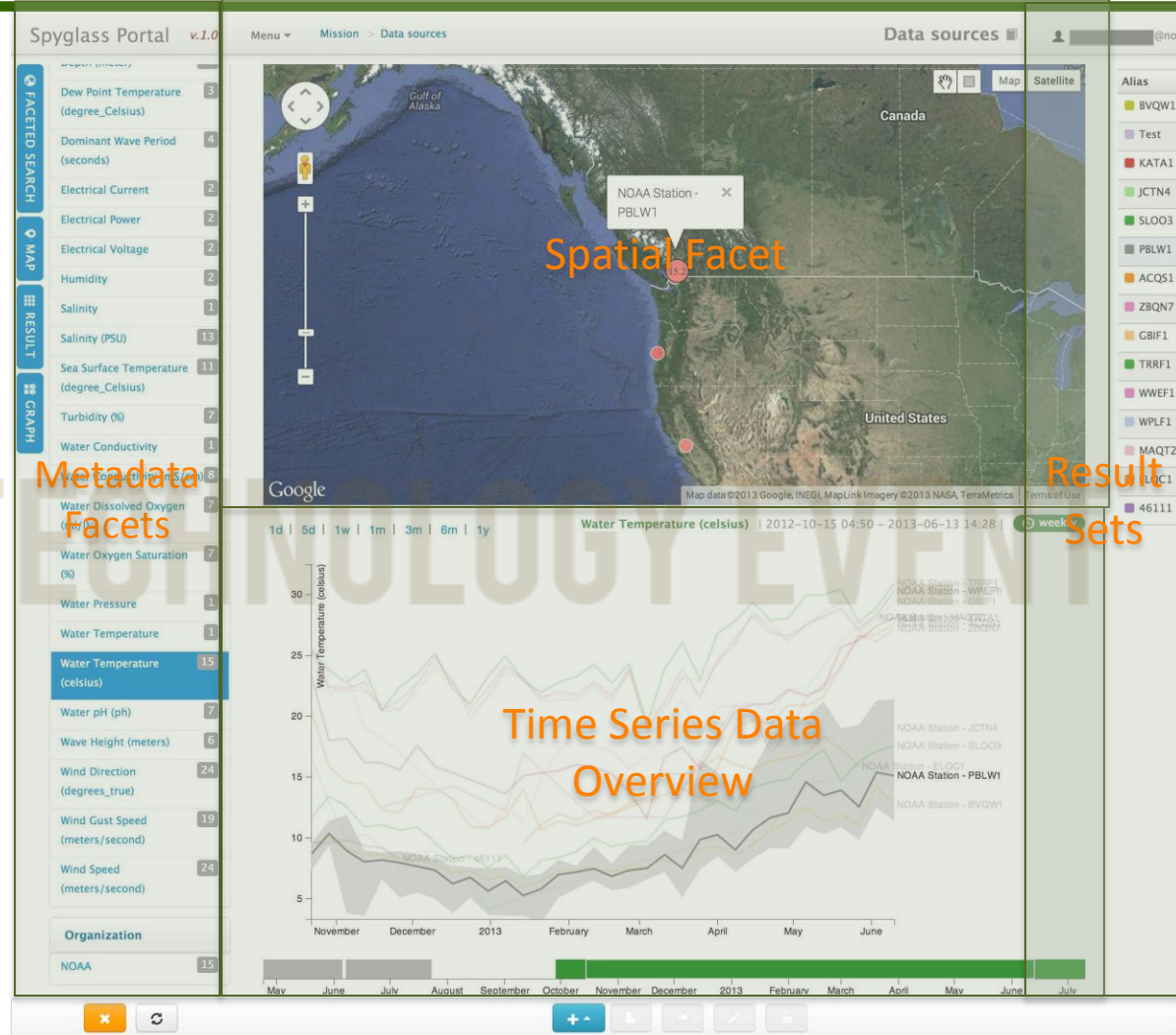
## Timeline of Events





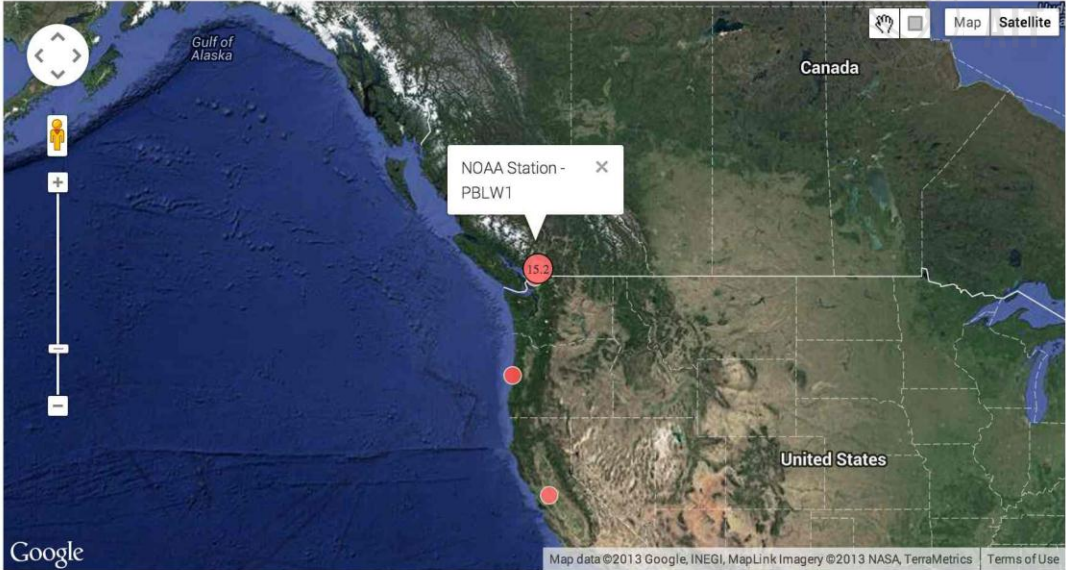
# Visualizing: Extended Network Data

- **Portal Data**
  - ESP Data
    - Organisms
    - Images
    - CTD, “Can”
    - Diagnostics
  - IOOS Data
    - Context

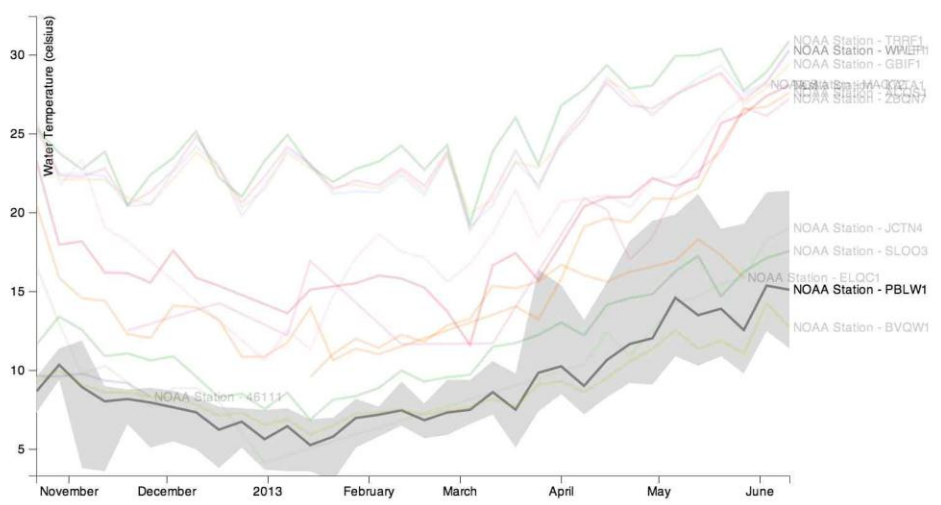


- Faceted Search
  - MAP
  - RESULT
  - GRAPH
- Dew Point Temperature (degree\_Celsius) 3
  - Dominant Wave Period (seconds) 4
  - Electrical Current 2
  - Electrical Power 2
  - Electrical Voltage 2
  - Humidity 2
  - Salinity 1
  - Salinity (PSU) 13
  - Sea Surface Temperature (degree\_Celsius) 11
  - Turbidity (%) 7
  - Water Conductivity 1
  - Water Conductivity (mS/cm) 8
  - Water Dissolved Oxygen (ml/l) 7
  - Water Oxygen Saturation (%) 7
  - Water Pressure 1
  - Water Temperature 1
  - Water Temperature (celsius) 15**
  - Water pH (ph) 7
  - Wave Height (meters) 6
  - Wind Direction (degrees\_true) 24
  - Wind Gust Speed (meters/second) 19
  - Wind Speed (meters/second) 24
- Organization**

NOAA 15



1d | 5d | 1w | 1m | 3m | 6m | 1y Water Temperature (celsius) | 2012-10-15 04:50 - 2013-06-13 14:28 | weekly



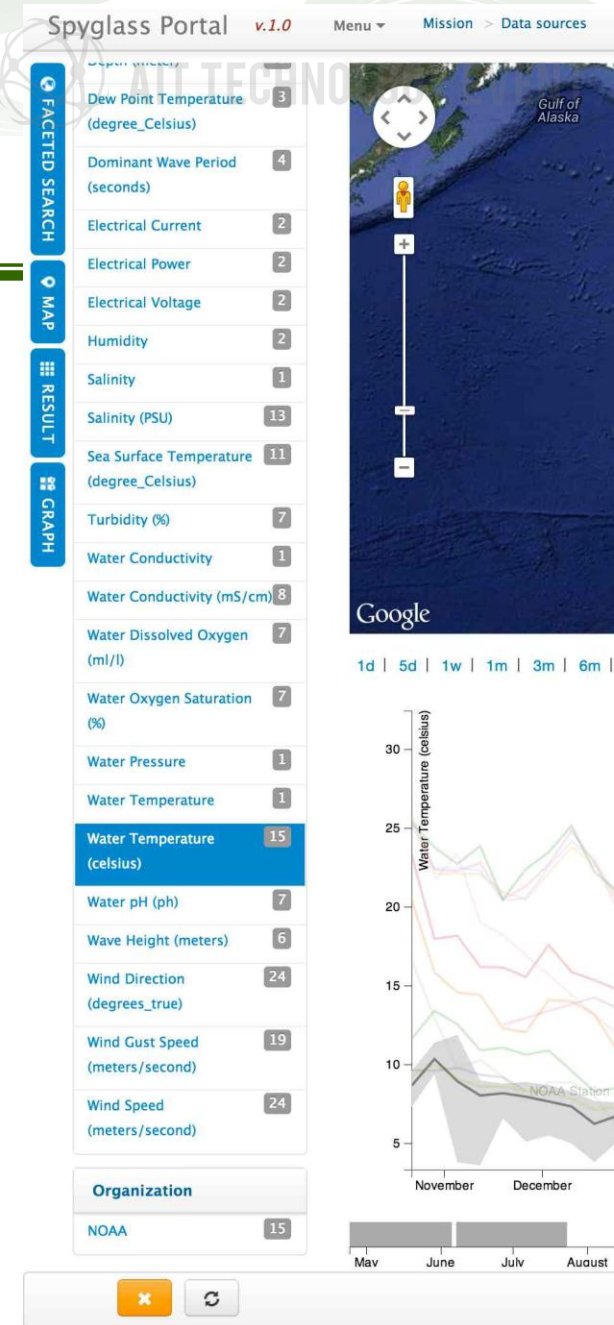
Alias	Status
BVQW1	Public
Test	Public
KATA1	Public
JCTN4	Public
SLOO3	Public
PBLW1	Public
ACQS1	Public
ZBQN7	Public
GBIF1	Public
TRRF1	Public
WWEF1	Public
WPLF1	Public
MAQT2	Public
ELQC1	Public
46111	Public



Dashboard

# Visualizing: Challenges

- **Web-based visualization**
  - Massive distributed data sets
  - Bandwidth constraints
  - Highly Interactive Graphics





# Summary: Seeing the Big Picture

	Challenges	Research & Technology	Applications
<b>Variety</b>	Structured: (Time Series) Unstructured: (Log Files)	Data and Metadata Standards Data Services & Discovery Data Integration Data Mining	Data Analysis Process Analysis Business Intelligence Decision Support Monitoring
<b>Volume</b>	Massive Data Sets	<b>Process Mining</b> <b>Faceted Search</b>	
<b>Velocity</b>	Real-Time Data	Distributed Databases Distributed Processing Distributed Architecture <b>Sensor Networks</b> Process Integration <b>Visualization Architecture</b>	





# Thank You



**AIT**  
Asian Institute of Technology

**Dr. Paul Janecek**

CEO, Think Blue Data

Visiting Faculty, Computer Science and Information Management

[paul@thinkbluedata.com](mailto:paul@thinkbluedata.com)