

H43B-1235: Berkeley Sensor Database (BSD)

An Implementation of CUAHSI's ODM for the

Keck HydroWatch Wireless Sensor Network

A System for Managing a Complex Sensor Observatory

Ginger Ogle¹, Collin Bode², and Inez Fung³

1. ginger@berkeley.edu, Earth & Planetary Sciences, UC Berkeley, Berkeley, CA, USA
2. collin@berkeley.edu, NCED, Integrative Biology, UC Berkeley, Berkeley, CA, USA
3. ifung@berkeley.edu, Earth & Planetary Sciences, UC Berkeley, Berkeley, CA, USA

ABSTRACT

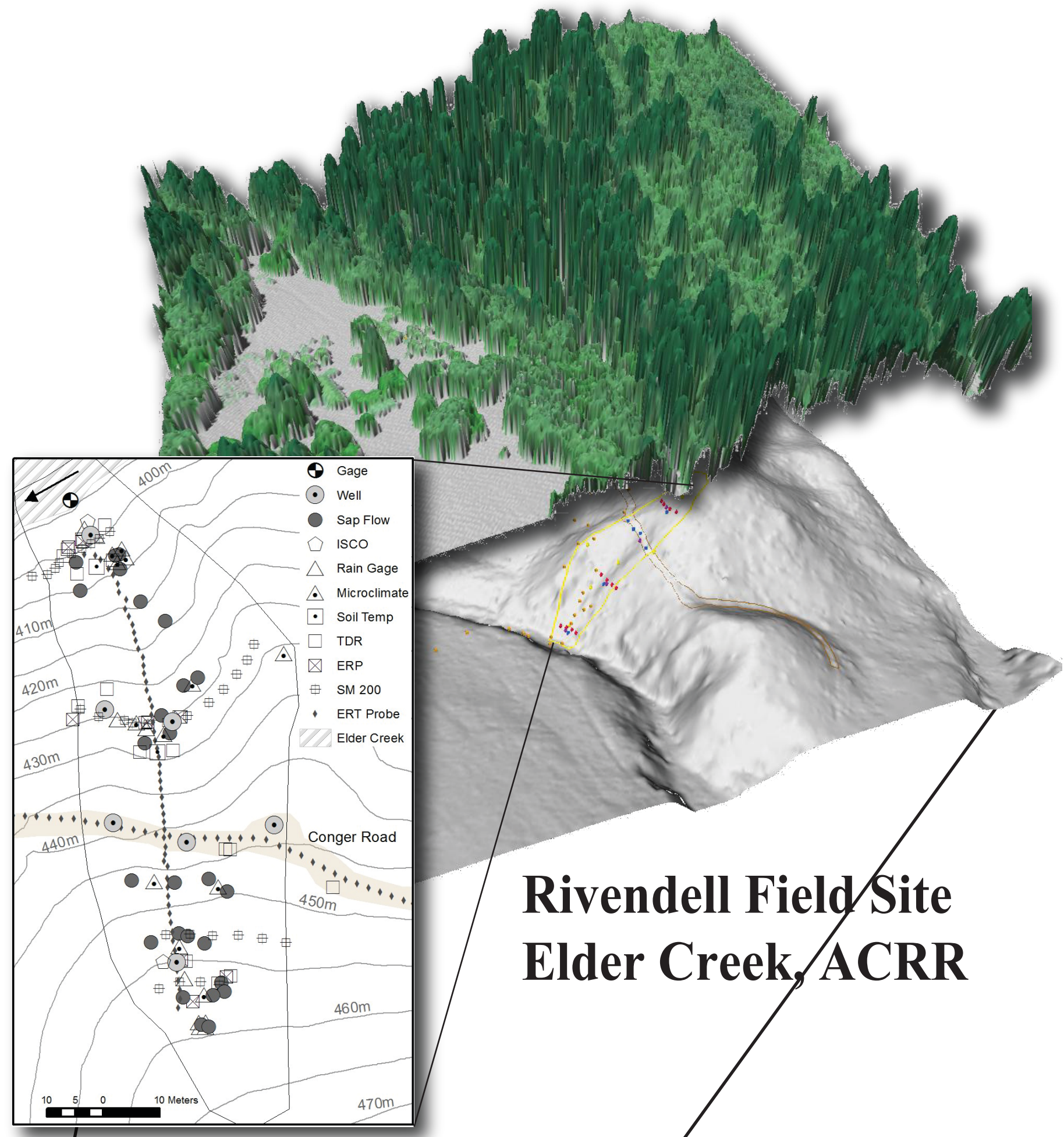
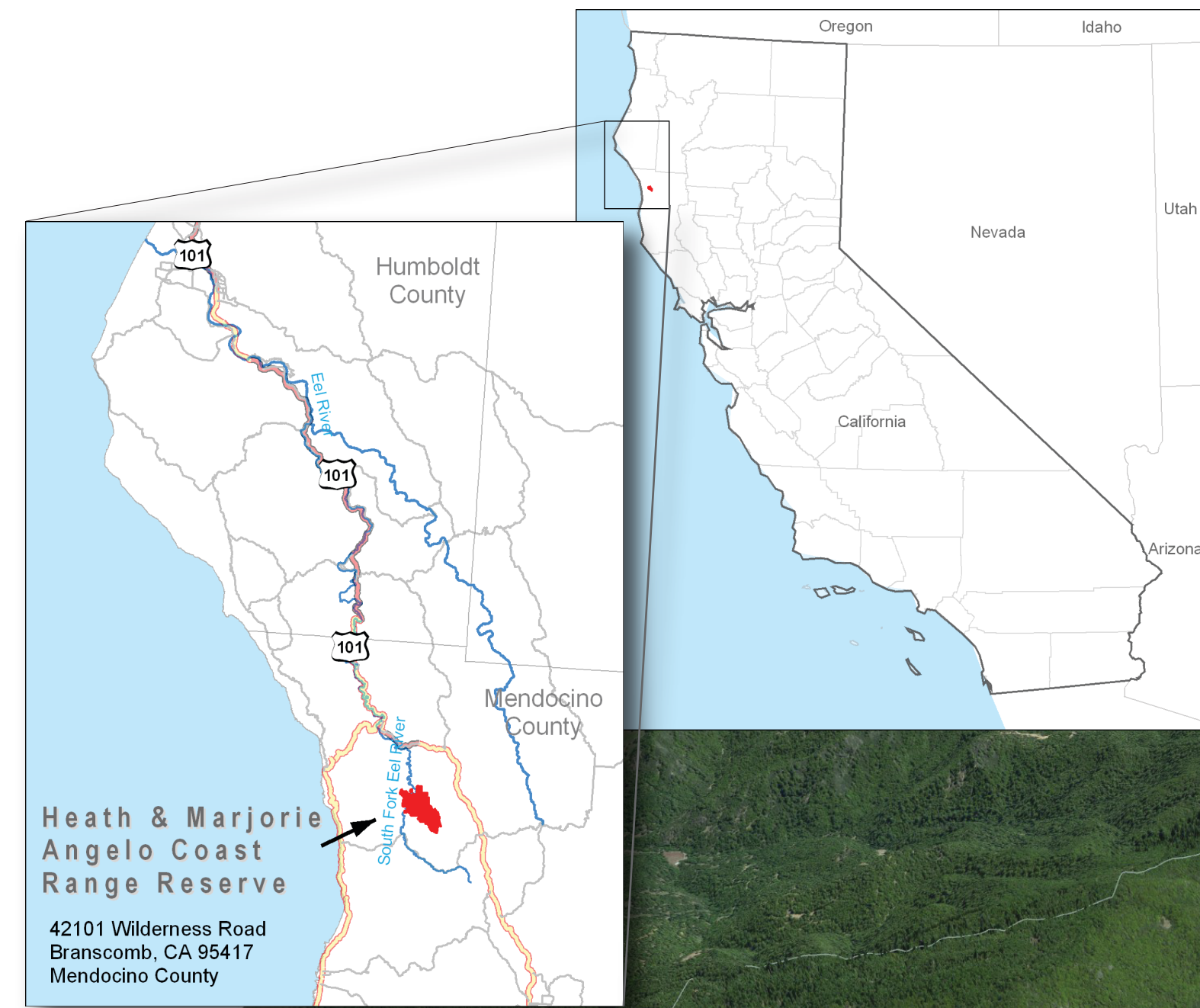
The **Keck HydroWatch Project** is a multidisciplinary project devoted to building a mechanistic understanding of the pathways, rates, and consequences of water movement through the soil, weathered bedrock, vegetation and return to atmosphere. Two key issues have been identified for intensive study: **fractured rock water** and **hydraulic lift & redistribution**. It is experimenting with novel techniques to monitor and trace water pathways through these mediums, including developing an intensive wireless sensor network.

The **Berkeley Sensor Database** was designed to manage the large volumes of heterogeneous data coming from this sensor network. This system is an open source superset of the Observations Data Model (ODM) developed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI).

Study Site

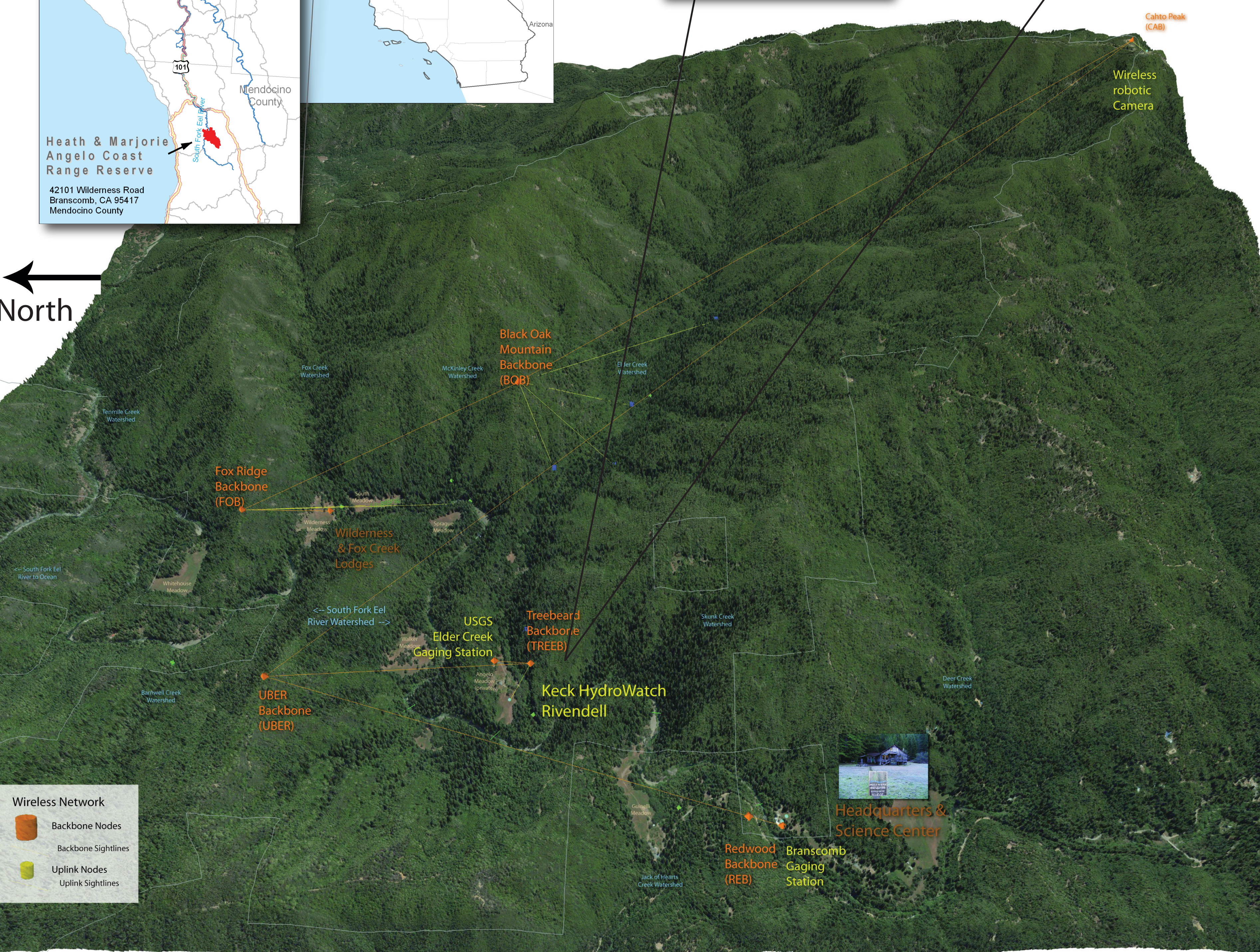
Angelo Coast Range Reserve

- Part of the University of California Natural Reserve System (35 reserves)
- Location: Mendocino County, California. 4 hour drive north of San Francisco
- Vegetation: One of the largest Douglas fir & coast redwood tracts remaining in CA.
- Oak-madrone forest and chaparral cover south facing slopes
- Geology: Franciscan complex argillite and greywacke sandstone
- Hydrology: 5km of the South Fork of the Eel River is within the Reserve
- Elder Creek Watershed: largest pristine watershed remaining in CA. Monitored by USGS as a benchmark for background purity of natural waters.



Rivendell Field Site
Elder Creek, ACRR

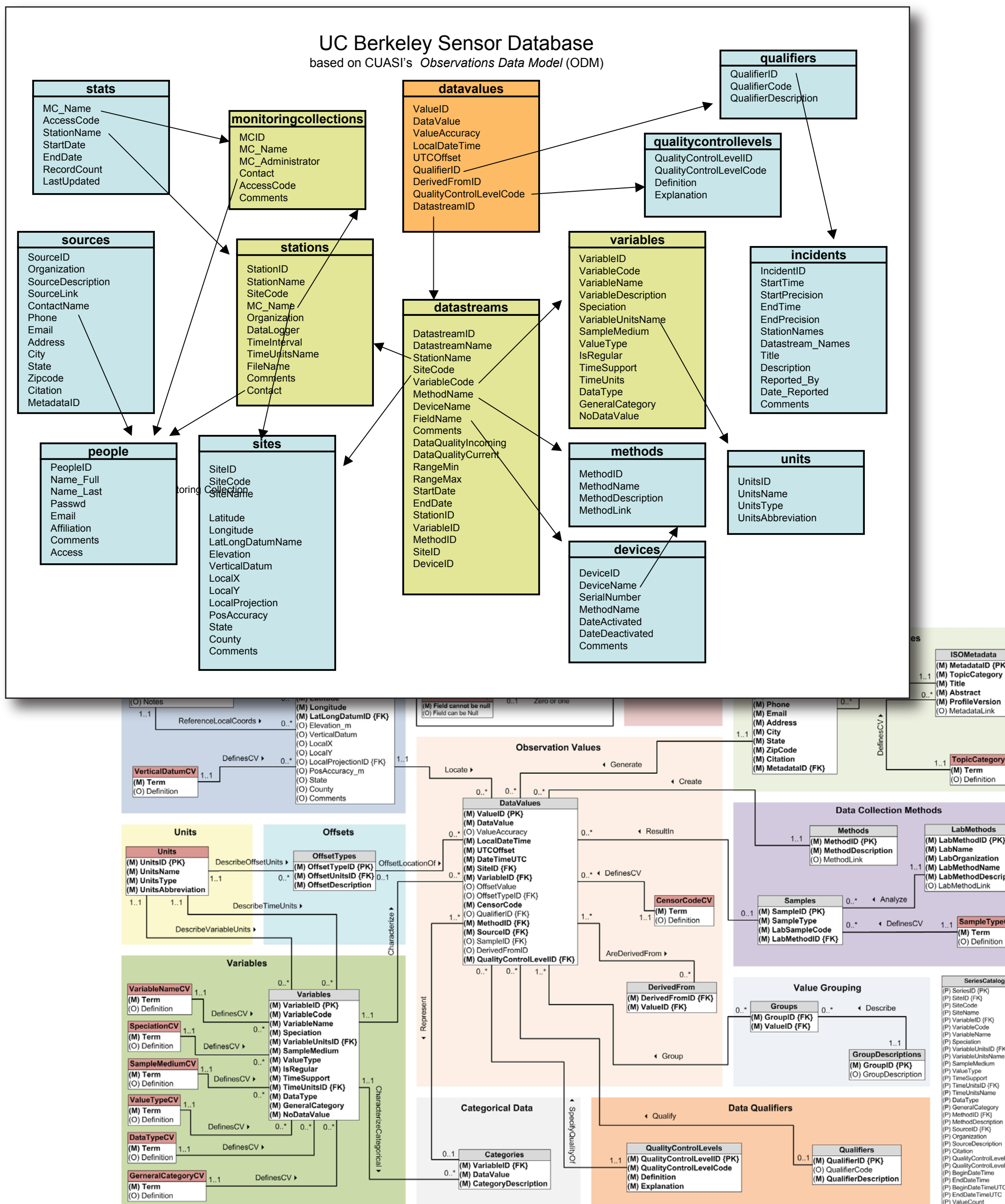
North



Why modify HIS?

1. Intensive experimental field site requires active management features:
- Allow for the movement and reuse of devices (datastreams)
 - Record changes in setup or accidents (incident reports)
 - Mechanisms to configure automated data loading (stations)
 - Access Control (Monitoring Collections)
2. Institutional mandate to develop on open source platforms
3. Keck HydroWatch Project had little access or support for Microsoft Platform:
- Microsoft Windows Server
 - SQL Server
 - MS trained IT personnel
5. Existing expertise in Open Source systems: Linux, Apache, MySQL, Perl (LAMP).
6. Building on existing codebase from systems developed for Berkeley Natural History Museum and NSF Digital Library Project.

BSD is compatible with HIS, but the web services have not yet been implemented on the LAMP platform.



Rivendell Field Site

Rivendell, the Keck field site, is a small unchanneled watershed within the Elder Creek watershed near its confluence with the South Fork Eel River. The north face has a slope of 32 degrees and is covered with mature Douglas Fir (*Pseudotsuga menziesii*). The south slope is dominated by Madrone (*Arbutus menziesii*). Soil is thin and the subsurface is primarily weathered argillite.

The study site is an intensive sensor observatory (most observatories are extensive) with over 1,000 sensors in a 200 x 100 meter area.

5 Douglas Firs (30 - 65 meters) are rigged for climbing and contain solar panels at the treetops, bringing power to the site.

Parameter Monitored	Campbell	Mote	Others*	Totals
Barometric Pressure mb	10			10
Air Temp C	30	29	32	91
Logger internal Temp c	11	29		40
Soil Temperature	69			69
Soil moisture TDR (dielectric)	100			100
Soil moisture ERP (resistance)	192			192
Soil Moisture SM200	30			30
Water Temp C	2	1	3	6
Gage Height meters	11	1	12	24
Turbidity NTU	1			1
Water Samples		2	2	4
Rainfall mm	12	1	13	26
Hail Cumulative mm	2		2	4
Hail Intensity mmh	2		2	4
Rel Humidity Perc	55	29	31	115
Snow Depth Meters	1			1
Total Solar Radiation W/m²	4	29	30**	63
PAR	29	35		64
Wind Direction Degrees	12	1	13	26
Wind Speed Avg MS	18		19	37
Sap Flow Velocity			154	154
Battery Voltage	24	29	1	54
Totals	575	174	290	1039

Sensors by Type



Wireless Networks

- Angelo Reserve has a multi-tiered wireless network.
- Tier 1: Backbone Network:
- Brings internet access into the Reserve.
 - Point-to-multi-point network
 - Eion VIP110-24 radios, 2.4ghz, VINES
 - Trees are used as towers.
- Tier 2: Campbell Scientific PakBus Network
- Connects the Rivendell stations to the Backbone network.
 - RF450 radios, 900mhz, pakbus
- Tier 3: Tertiary Networks
- Wifi: 802.11, Reserve housing and select hotspots on in the field.
 - Provides internet access for researchers with laptops.
 - Motes: experimental mesh networking motes with microclimate data. Prof. David Culler, CS, UCBC.

Statistics

<http://sensor.berkeley.edu>

42,536,000 Measurements

178,199 Measurements/Day

1,294 Datastreams

661 Instruments

68 Sites

29 Researchers

(as of Dec 8, 2010)

Database Schema Modifications

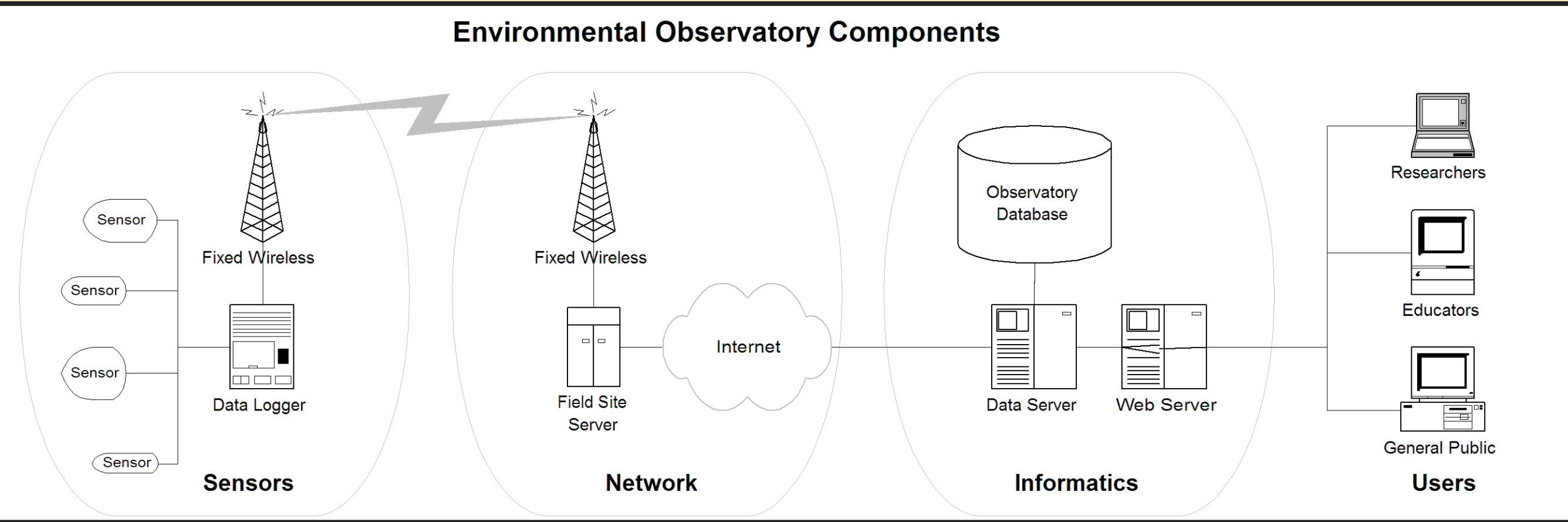
- Project Requirements:
- Optimize web-based queries on a large DB
 - Optimize data loading
 - Build in tighter quality control & access control
 - Adapt model for broader range of sensors (e.g. sap flow)
- Implementation:
- Create new devices, datastreams, sites, etc.
 - Record "incidents" and flag data
 - Edit all ODM metadata (methods, variables, sites, etc.)
 - Automated email alerts for error conditions
 - Automated alerts on schema modification
 - Manage access permissions for participants
 - Monitor equipment & data acquisition
 - View up-to-the-minute stats (count, timestamps)

- Implementation:
- 1. Modify some tables
 - Omit some fields (eg, datavalues 17 >> 9 fields)
 - Pare down data types (double >> tinyint)
 - Repurpose tables (method = device type)
 - Add datastream keys to some tables
 - 2. Drop some tables
 - Use app-level structures instead (e.g. Categories)
 - Delay implementation for some tables (e.g. catalog)
 - 3. Add new tables
 - Optimize queries: datastreams ("superkey")
 - Access control: people, monitoring collections
 - Quality control: incidents
 - Data mgt: datastreams, devices, stations, stats

Access Control

- Project Requirements:
- Restrict access to research-level data
 - Provide public access for public datasets
- Implementation:
- User accounts & passwords (including "guest")
 - Access levels for each person
 - Access levels for each monitoring collection
 - Auto alerts to admin when metadata is edited, data is added

Data Workflow



Networks: Workflow

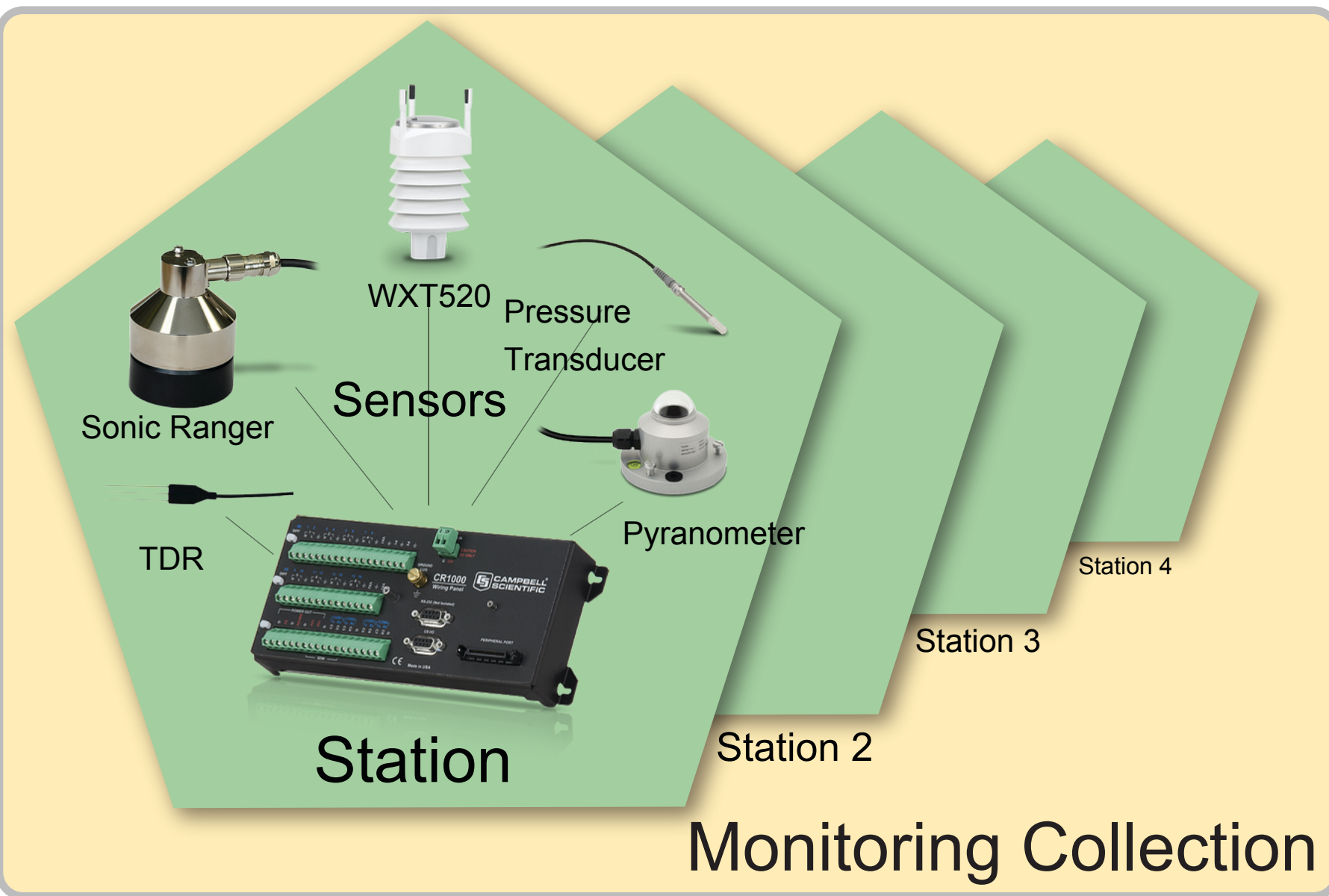
- Networks connect distributed stations with Reserve HQ.
- LoggerNet Software (Campbell Scientific) is used to pull data from stations.
- Remote reconfiguration & status monitoring performed through LoggerNet.
- Caching Server stores data files in comma separated value (csv). Files are rsynced every 30 minutes to BSD server.



Integrating research from field to lab to model makes for happy scientists.

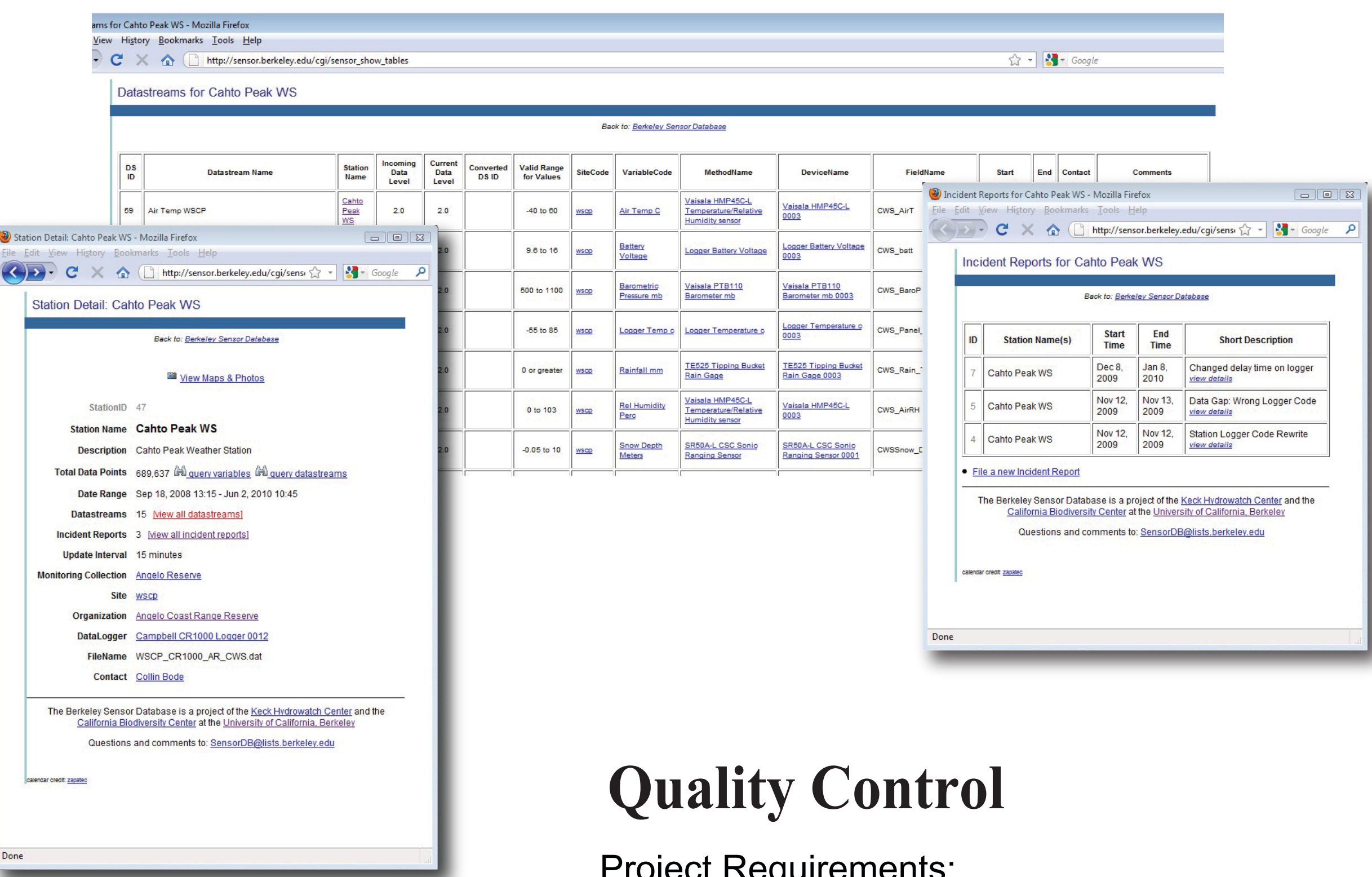
Informatics: Data Loader

- The Data Loader runs every 30 minutes for a variety of logger types (CS CRbasic, Postgres/Motes, ICT sap flow custom loggers), as well as auto loading for hand measurements uploaded via web.
- Checks for new data arriving from the field
 - Checks for new devices/logger configs
 - Performs sanity check: flags out-of-range data, emails alerts
 - Assigns data quality level
 - Converts raw data to geophysical units as needed
 - Inserts new records



Management Functions

- Project Requirements:
- Web-based management tool
 - Remote device configuration
- Implementation:
- Create new devices, datastreams, sites, etc.
 - Record "incidents" and flag data
 - Edit all ODM metadata (methods, variables, sites, etc.)
 - Automated email alerts for error conditions
 - Automated alerts on schema modification
 - Manage access permissions for participants
 - Monitor equipment & data acquisition
 - View up-to-the-minute stats (count, timestamps)



Quality Control

- Project Requirements:
- Store all versions of data (raw, converted, derived, corrected)
 - Sanity check on incoming data
 - Allow users to flag data
 - Show flags & annotations on all data
- Implementation:
- Data Quality Levels (odm.qualitycontrollevels) based on NASA EOS
 - Data Qualifiers (odm.qualifiers) sanity check + incidents
 - Loader performs sanity check & flags data (P,U,VB,VE,X)
 - Loader adds Data Level for raw, converted, derived
 - Incident Reports: users can flag data
 - Data query/download: view flags, exclude flagged data

Data Qualifiers

Data Flag	Description
P	Passed sanity check
U	Unchecked value, no sanity check
NV	No value recorded (NaN)
VB	Value below device minimum range
VE	Value exceeds device minimum range
X	Invalid Data: equipment malfunction

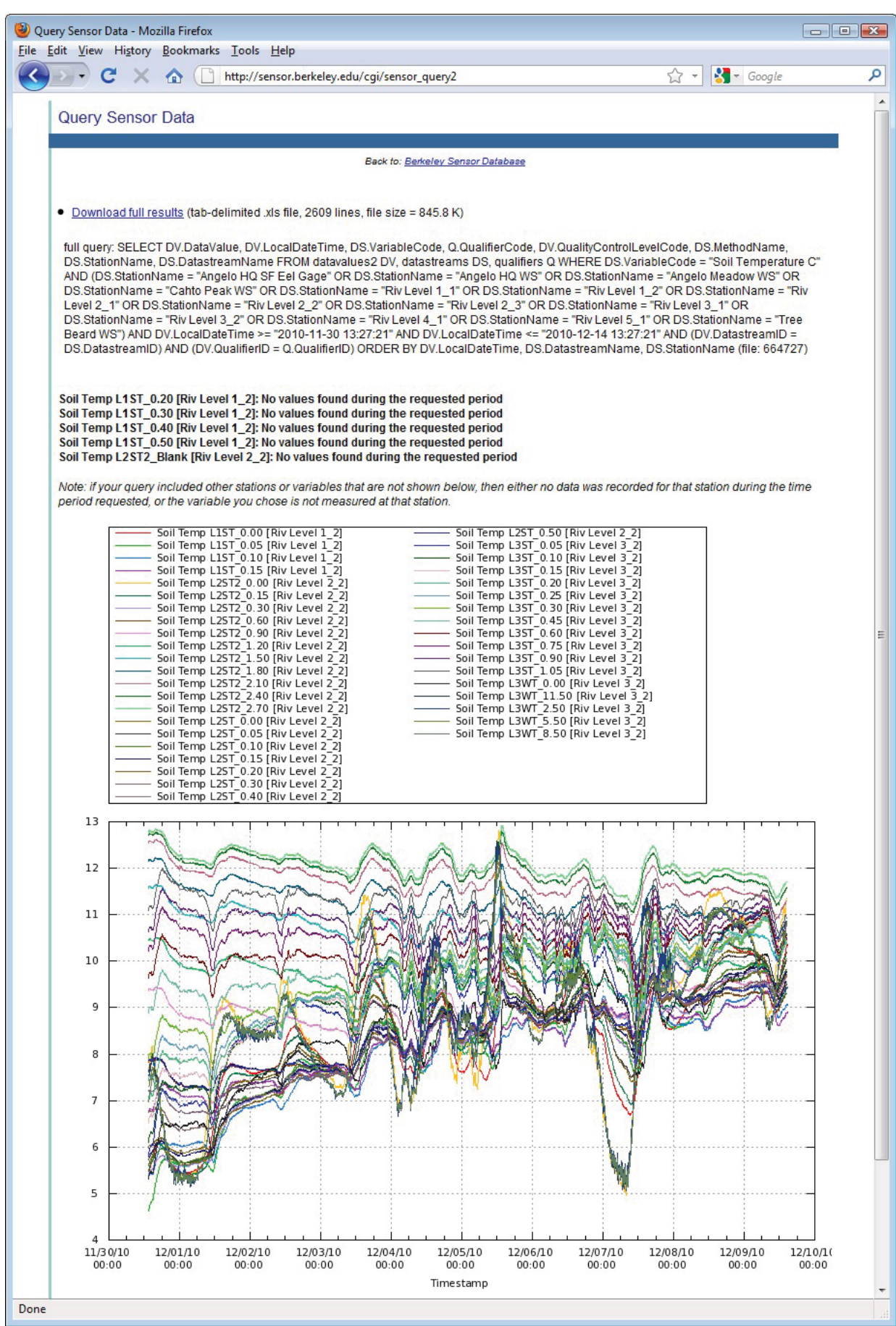
Data Quality Levels

Level	Definition
0	Instrument Data
1A	Instrument Data + Metadata
1B	Instrument Data in Sensor Units + Metadata
2	Geophysical Units
3	Geophysical Units + Space/Time Uniformly Scaled
4	Derived Data, from Multiple Variables or from a Model

Conceptual Design

- **Monitoring Collection:** A conceptual grouping of stations, usually being used by a specific research project or organization.
- **Station:** One datalogger with sensor devices attached, such as a weather station. The Station is also used to define the output data file which is pulled from the logger.
- **Device:** A specific instance of a method that is installed. A device can be decommissioned, or moved to another location.
- **Datastream:** The "data flow" of one type of data from a device of a particular method at a particular location.

Type of Access	Requirements	Data Format
Web Access	Web browser & valid project login	Graph & CSV download
Direct MySQL Access	Knowledge of SQL	SQL output
Matlab	Matlab MySQL connector or Mathworks Database Toolbox	Matlab matrix format
Bulk Data Dumps	Valid project login	Zipped, CSV, weekly refresh



Incident Report ID#1

Title: Bear damage
Station Name(s):
Riv Level 1_1, Level 1_2,
Riv Level 2_1, Level 2_2,
Riv Level 3_1, Level 3_2,
Riv Level 4_1
Datastream(s):
resistance probes,
rain collector
wells
Data Flagged:
ERF: 290,304 values
Well 6: 3,024 values
Well 7: 3,024 values
Start Time: 2009-09-24 (+/- 1 month)
End Time: 2009-10-24 (+/- 1 day)
Reported By: Bill Dietrich (2009-10-26 00:00:00)
Description: We visited the site this past weekend. Unfortunately a bear did some damage recently. It systematically knocked over and chewed on everyone of the plastic bag covered resistance probes. They were each snapped at the soil boundary. I haven't had a chance to check to see when this might have happened according to the online record. The bear also knocked over the rainfall collector (James says he has found teeth marks in it before). It broke the extended plastic tube on well 7 and the wooden stake holding up the plastic tarp, and it knocked off the cap of well 6. We didn't detect other damage-- but there could be some. I encourage you all to check your favorite on line device. We put the cap back on 6, propped up well 7 tube and put back a stake, and reset up the rain collector.



Support & Collaborations

W.M. Keck Foundation Science & Engineering Research Grant, 2005
National Center for Earth-surface Dynamics (NCED)
University of California Natural Reserve System
Berkeley Institute of the Environment
Consortium of Universities for the Advancement of Hydrologic Sciences, Inc.
Berkeley Natural History Museums (BNHM)

