Monitoring the Monitors: Assessing and Visualizing the State-of-Health of Earthscope's Transportable Array

Newman, R. L. <rlnewman@ucsd.edu>[1], Lindquist, K.G.[2], Foley, S.[1], Davis, G. A.[1], Vernon, F. L.[1], Eakins, J.[1] and Astiz, L.[1]
[1]- Scripps Institution of Oceanography, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0225 [2]- Linquist Consulting, 59 College Rd. Suite #7, Fairbanks, AK 99701

IN23A-1213

1. ABSTRACT

The Earthscope Transportable Array has increased in size over the last three years to over 200 broadband seismic stations. Assessing the state-of-health of the station equipment, collecting data transfer metadata, and providing this information to analysts, station engineers, administrative staff, researchers and the public is the responsibility of the Array Network Facility (ANF).

Various interconnected software packages (including the Antelope Environmental Monitoring System, RRD, GMT, MATLAB, Nagios and Flickr) build data products in near real-time that are organized and integrated into the ANF website using PHP (a web-based scripting language embedded in HTML). These metadata and data products are readily accessible via the world-wide-web at http://anf.ucsd.edu, where multiple web-based tools have been developed to display and visualize these products.

2. THE PROBLEM

The Transportable Array (TA) is one component of the Earthscope USArray project. It consists of 400 broadband seismic stations delivering data in real time to the Array Network Facility (ANF), housed at the Scripps Institution of Oceanography. In turn the data is sent to IRIS for archiving and also stored offsite in a third location. The network is currently ramping up to full capacity. As of December 2, 2006 there are 293 stations operational.

The ANF is responsible for data collection and its quality control, and State-Of-Health (SOH) monitoring of the stations. The latter includes the seismic sensor, datalogger, vault conditions and data transfer rates from the station to the archives (see Box 4).
The high sample rate of sensors (40 samples per second) and the fact that data is delivered in quasi-real-time demands a robust monitoring system that is accessible by station engineers, ANF staff, collaborators and end-users of the data.

3. THE SOLUTION: WEB-BASED ANALYSIS

To facilitate monitoring of such a large network, and to allow as diverse an audience as possible, we have developed a suite of online-web based tools (see Box 6 and 7) available from the public ANF website, with further administrative tools located over an encrypted/authenticated secure connection. Access to the secure administrative area is limited to station engineers and analysts.

The development of these tools allows any end-user to choose a station(s) of interest and request and/or observe SOH data for any number of components. In addition to monitoring tools that we have written, we have also leveraged pre-existing open-source and proprietary tools. These various tools are the focus of this poster.

4. COMPONENTS TO BE MONITORED

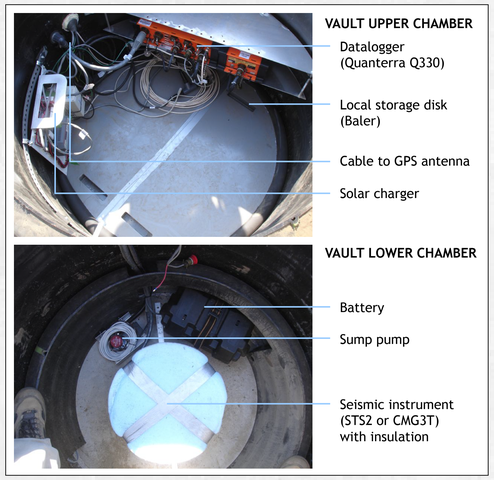


Figure 1. Upper and lower view of TA vault, with visible components labeled

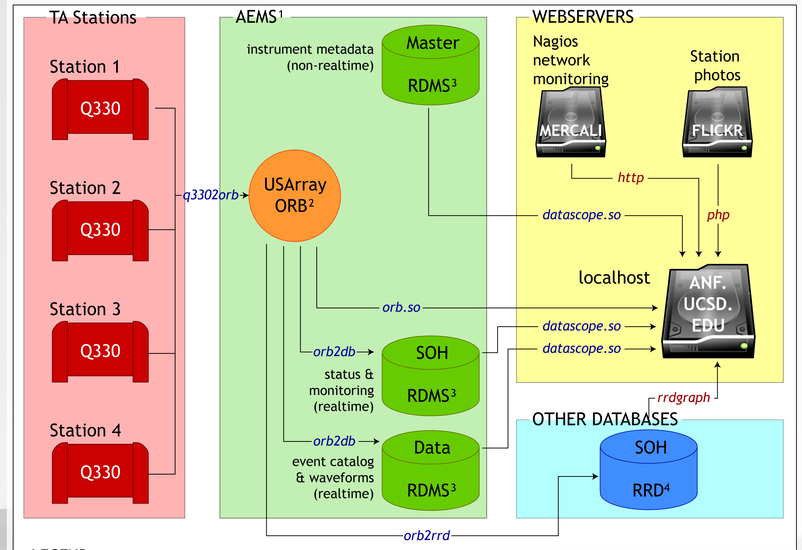
Each TA station has a variety of environmental sensors that are monitored.

Data packets along with log messages, status messages, command responses and debug messages are written to a set of output ORBs.

A large set of status channels can be generated as regular ORB waveform packets.

The data is sent to an ORB using q3302orb.

5. STATE-OF-HEALTH MONITORING SCHEMATIC



LEGEND
[1] - Antelope Environmental Monitoring System
[2] - Object Ring Buffer
[3] - Relational Database Management System
[4] - Round Robin Database
php - Data transfer mechanism
q3302orb - AEMS data transfer mechanism

Figure 2. Schematic diagram of how State-Of-Health (SOH) information travels from the datalogger in the field to the ANF website. Data is stored in different database types for different purposes. Mulitple web servers house different tools and applications. A commercial webserver, www.flickr.com, is used as a photograph repository and allows station engineers to upload photos directly from their laptops in the field to a secure account. Selected photographs are pulled from the flickr database and displayed on the ANF website.

6. USER INTERFACE TO STATE-OF-HEALTH MONITORING

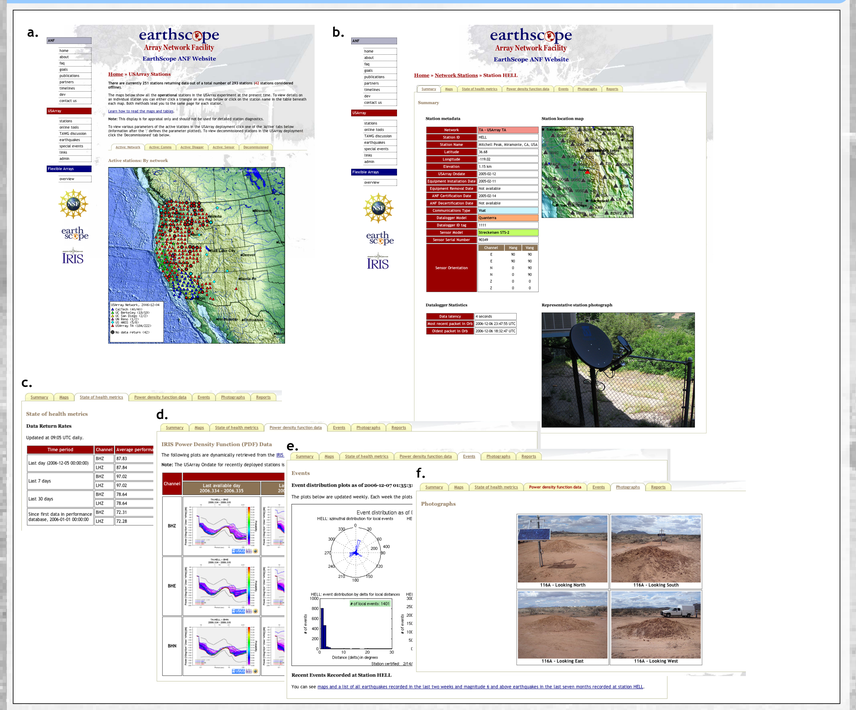


Figure 3. Stations are displayed on regional maps (**a**) of network, communication type, datalogger model, and sensor type. Decommissioned stations are also displayed on a map. Stations enclosed with a gray circle are considered offline (data latency > 4 hrs.) Legend shows stations in each category (online/total ratio shown). Clicking an icon on the map, or table entry, returns a detailed page for the respective station (**b**). Tabs display multiple SOH information,including data return rates (**c**), power density function (PDF) plots (**d**), events recorded by the station (**e**), and photographs of the station (**f**). The state-of-health metrics tab also shows the last hour of various state of health parameters in Round Robin database Archives (RRAs). See Box 7 for details.

7. DETAILED ANALYTICAL TOOL: ON-THE-FLY QUERIES AND RESULTS WITH RRD

Round Robin Database (RRD) tool is an industry standard system to store and display time-series data. It's origins lie in the network monitoring world, however it can be used to display any time-series data. We have developed a method of populating a RRD Archive (RRA) with data from an ORB, called orb2rrd, that runs as part of a real-time system. Whenever a new SOH status packet arrives in the ORB the packet is opened and the contents disseminated to a suite of RRD archives.

Database size is determined at creation time. The databases never increase in size, even with high sample rates. This is because the RRAs are not linear databases - it helps to think of RRA's as the perimeter of a circle: new data gets added along the perimeter. When new data reaches the starting point, it overwrites existing data. Different levels of data decimation are defined at creation time, allowing long time periods to be displayed quickly and with accuracy.

For each station we assign a single variable (such as vault temperature) to its own archive. In this way we can retrieve data from different archives and plot them in a variety of ways, either singularly or in combination with other variables.

The variables monitored cover eight broad areas considered to be the most useful to analysts and station engineers in diagnosing state-of-health problems: clock operations, vault conditions, sump pump disposition, mass position, data input and output, efficiency, datalogger bytes read and written, and communication link cycles and reboots.

Defined time periods available for plotting are: last hour, last day, last week, last month and last year. The graphs are created on request from the webpage. The small size of each archive allows rapid production of each graph, essentially in 'real-time'.

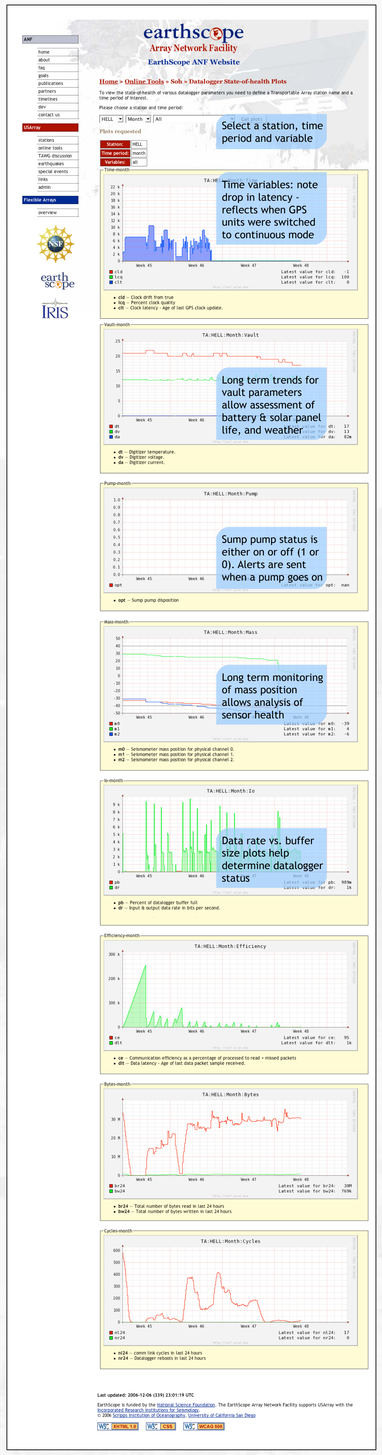


Figure 4. Screen dump of the SOH plotting webpage that uses the RRD database and graphing capabilites. Variables to be plotted are defined in form fields. Graphs are generated after hitting the 'Get plots' button. Highlights for each graph are shown in blue.

8. OTHER TOOLS

The Antelope suite of tools comes with several monitoring applications, such as a real-time waveform trace display (**orbmonrtd**) and a datalogger monitor (**dlmon**). Both these X-windows based tools can be modified to allow image dumps of their displays. We typically run these applications using Xvnc and display the dumped images on the ANF website. These tools are currently being replicated in a pure web-based programming environment, with the focus on development of webdlmon.

9. FUTURE PLANS

As the Transportable Array continues to grow, more online analytical tools need to be developed to allow timely notification, analysis and decision making in real-time.

Increased station numbers also means more time is required to run network wide analysis tools. We have run into problems with webserver load due to high traffic rates and database queries. To improve this we are currently developing new tools and modularizing and refining our current toolset.