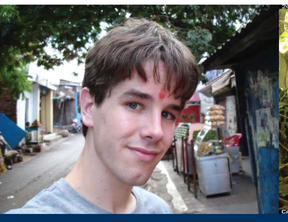
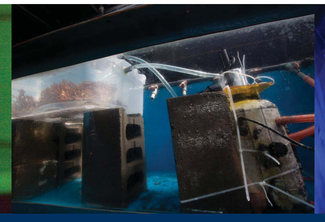
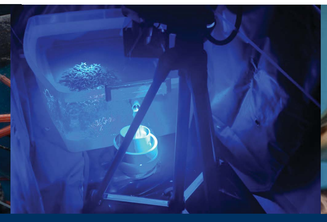
As the study of natural phenomenal is increasingly being tied to high tech engineering, scientists are finding themselves faced with new challenges in conducting their research. Large amounts of data from varied and numerous sensors must be handled in real time. These sensor networks are distributed and complex, and are comprised of a heterogeneous set of instruments. Commonly incorporated sensors include everything from high definition streaming video camera, to seismic activity detectors, to geospatial tracking systems. Streaming data must be available in a timely manner, and must be visualized in a useful manner. The Open Source DataTurbine (OSDT) Initiative was launched in October 2007 with a two-year grant from the National Science Foundation Office of Cyberinfrastructure (award #OCI-0722067) to address these challenges through the publication, enhancement, and promotion of the DataTurbine streaming data middleware (www.dataturbine.org).

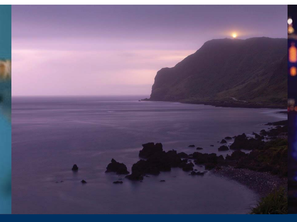
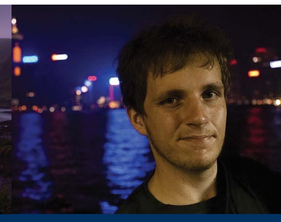
DATA TURBINE

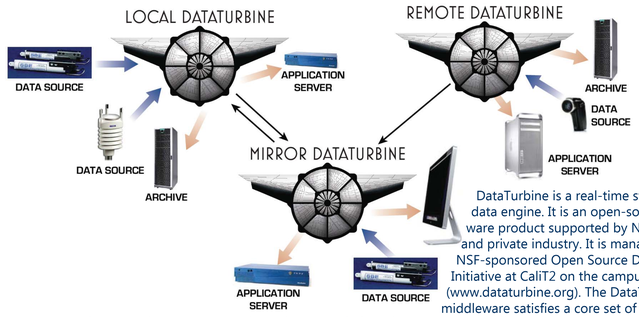
The OSDT initiative has been working in conjunction with the Pacific Rim Undergraduate Experiences program (PRIME), started by PRAGMA in 2004, to help facilitate undergraduate research experiences. Students have been working at home and abroad to help develop DataTurbine based systems to meet the needs of collaborators around the world. Most recently in Summer 2009, UCSD undergraduates Brian McMahon and Michael Nekrasov were sent to sites in India and Taiwan to deploy DataTurbine systems. They have since continued research work with PRAGMA and the OSDT initiative, and are helping to coordinate sending even more students abroad during the upcoming summer.







DataTurbine is a real-time streaming data engine. It is an open-source middleware product supported by NSF,NASA, and private industry. It is managed by the NSF-sponsored Open Source DataTurbine Initiative at CaliT2 on the campus of UCSD (www.dataturbine.org). The DataTurbine middleware satisfies a core set of infrastructure requirements that are common in environmental observing systems, including reliable data transport, a framework for integrating heterogeneous instruments, and a comprehensive suite of services for data management, routing, synchronization, monitoring, and visualization. DataTurbine provides a programming abstraction in which data streams are first-class objects with a rich set of services for creating and managing streaming data applications. From the perspective of distributed systems, the DataTurbine middleware is a "black box" to which applications and devices send and receive data. DataTurbine handles all data management operations between data sources and sinks, including reliable transport, routing, scheduling, and security. DataTurbine accomplishes this through the innovative use of flexible network bus objects combined with memory and file-based ring buffers. Network bus objects perform data stream multiplexing and routing. Ring buffers provide tunable persistent storage at key network nodes to facilitate reliable data transport. DataTurbine has rigorously tested in a wide variety of demanding applications and has proven to be robust, efficient, scalable, and easy to use. It has been deployed on a variety of compute platforms, from cell phones to supercomputers. DataTurbine is published under the Apache 2.0 open source license and is available for download at http://code.google.com/p/dataturbine/.

MoveBank
Animal Tracking

The MoveBank project is a open community with the goal of organizing and making available data related to the remote observation of animals in their natural habitat. Their website (www.movebank.org) has data sets from more than one hundred studies around the globe, all freely accessible to researchers and students. The OSDT initiative has been working with MoveBank, using the data from the researchers at Barrio Colorado Island (BCI) in Panama, to develop technology to facilitate the visualization and acquisition of geospatial data for animal tracking. The collaborators in this project include Dr. Tony Fountain and Sameer Tilak from CalIT2 at UCSD, Dr. Roland Kays from the New York State Museum, and Brian McMahon, an undergraduate student at UCSD.

The system in place at BCI involves tracking collars on ocelots, monkeys and agoutis sending data to a network of local receiver towers, as the density of the jungle canopy makes getting an accurate GPS reading impossible. Transmitters are also being placed on seeds to learn more about the food caching habits of agoutis so as to better understand their relationship with their habitat. Movement data from these animals is being integrated into DataTurbine as a method of efficiently streaming information to researchers. Current development includes a map based visualization system designed to natively run alongside DataTurbine to allow for real time, meaningful analysis of position data.

PRIME @ UoH
Tsunami Early Warning System

One of the greatest natural threats faced by populations in many coastal areas is the destructive force of tsunami waves. In the wake of the widespread destruction of the 2004 Indian Ocean Tsunami, the world has considered protecting people from these hazardous waves more important than ever before. As prevention of tsunamis is not possible, the key to minimizing loss of life in these situations is early detection and response. With an early enough evacuation notice, the potential devastation of the wave can be greatly mitigated.

In an effort to combat this global issue, UCSD and CalIT2 are participating in facilitating undergraduate research experiences in India in partnership with the University of Hyderabad. In Summer 2009, UCSD computer science undergraduate student Brian McMahon was sent to Hyderabad to work with the Indian National Centre for Ocean Information Services (INCOIS) on the development of a tsunami early warning system. The overseeing collaborators were Dr. Tony Fountain and Sameer Tilak from the OSDT initiative at CalIT2, and Dr. Arun Argawal at the University of Hyderabad. The collaboration was funded by the NSF and was organized by PRAGMA's PRIME program.

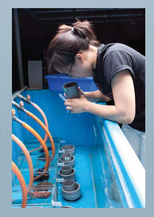
The early warning system at INCOIS utilizes real time data from a large global network of sensors, including earthquake monitoring facilities, ocean floor pressure sensors, and coastal tide gauges. These sensors provide data in real time, which is used to trigger alerts and compare with pre-calculated models to predict the time frame and area effect of a tsunami wave. DataTurbine is being used to more efficiently process this data, with the goal of helping facilitate an automated decision support system for the generation of tsunami alerts. The end goal of developing this system is to improve the speed and accuracy of alerting disaster preparedness organizations around India so that as many lives as possible can be saved.

PRIME @ NMMBA
Coral Monitoring

Coral reefs are a diverse and dynamic environment, and they continue to flourish in activity even after the sun has set. Diving takes large amounts of time and can be treacherous during night time. Because of these issues, it is beneficial to setup long term underwater surveillance systems to monitor the behavior of the reef even during times the human eye can not. Data collected by sensors on the reef can often give indicators that an event of interest is occurring and so can be used to give an early warning to scientists that more direct observations need to be conducted.

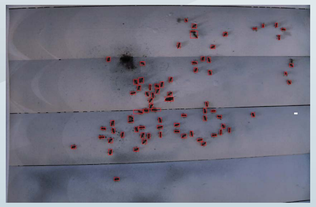
During the summer of 2009, Michael Nekrasov, an undergraduate student at UCSD majoring in Computer Science and Mathematics, was sent to the National Museum of Marine Biology and Aquarium in Taiwan as part of UCSD's PRIME program. The project was a joint collaboration between CALIT2 and Scripps Institute of Oceanography (SIO) in San Diego, and the Taiwanese National Center for High-performance Computing (NCHC) and Museum of Marine Biology and Aquarium (NMMBA). The principle collaborators were Dr. Tony Fountain, the PI for the Open Source DataTurbine Initiative at UCSD CALIT2, Dr. Dimitri Deheyn and his PHD student Melissa Roth, researchers at the Marine Biology Research Division at SIO , Dr. Fang-Pang Lin , a researcher in information technology at NCHC and coral biologist Dr. Tony Fan from NMMBA.

The project used an excitation light and blocking filter to excite the fluorescent protein in Seriatopora Hystrix to capture the fluorescence on video. The fluorescent light allowed hard to see coral features to be visible at night, and was also used to study coral spawning events (which are normally difficult to record since they occur at night and involve coral larvae millimeters in diameter). The video was streamed in real time using Data Turbine to an image recognition program, developed by PRIME student Robert Chen at UCSD, to identify and count the coral larvae and store the results back into Data Turbine in real time.



TFRI
Bee Counting

With the rapid development and industrialization happening around the world, environmental changes are occurring at a rapid rate. It is thus important to develop technologies to aid scientists in the study of environmental changes. Bees for example are an important indicator to the relative state of an ecosystem.



Building upon the coral monitoring project from 2009, a new collaboration was set up with the Taiwan Forestry Research Institute (TFRI) to tap into the infrastructure of cameras in place to monitor various environmental activities like bee interaction. This system uses the same type of framework researched at NMMBA, Taiwan to provide researchers with a set of tools to automatically count changes in bee populations, and to automatically monitor and record bee and wasp interaction.

Although the research is application driven towards providing TFRI with an automated system for bee counting, the framework being developed will have broad application to counting and registering environmental events in many fields, including coral biology. There are currently plans in place to send several students to TFRI for summer 2010 to widen the breadth of this collaboration.

This project is a collaboration between. Dr. Tony Fountain, the PI for the Open Source DataTurbine Initiative at UCSD CALIT2, Dr. Sheng-Shan Lu and Dr. Kevin Chen at TFRI, and Michael Nekrasov, an undergraduate in Computer Science and Mathematics at UCSD.

primary url: www.dataturbine.org

open source: code.google.com/p/dataturbine