



## **A Data Scheduling and Management Infrastructure for the TEAM Network**

S. Andelman (1), C. Baru (2), S. Chandra (2), E. Fegraus (1), K. Lin (2), and R. Unwin (2)

(1) TEAM Network, Center for Applied Biodiversity Science, Conservation International, (2) San Diego Supercomputer Center, UCSD, Science R&D, La Jolla, United States (chandras@sdc.edu)

The objective of the Tropical Ecology Assessment and Monitoring Network ([www.teamnetwork.org](http://www.teamnetwork.org)) is “To generate real time data for monitoring long-term trends in tropical biodiversity through a global network of TEAM sites (i.e. field stations in tropical forests), providing an early warning system on the status of biodiversity to effectively guide conservation action”. To achieve this, the TEAM Network operates by collecting data via standardized protocols at TEAM Sites. The standardized TEAM protocols include the Climate, Vegetation and Terrestrial Vertebrate Protocols. Some sites also implement additional protocols. There are currently 7 TEAM Sites with plans to grow the network to 15 by June 30, 2009 and 50 TEAM Sites by the end of 2010.

### Climate Protocol

The Climate Protocol entails the collection of climate data via meteorological stations located at the TEAM Sites. This includes information such as precipitation, temperature, wind direction and strength and various solar radiation measurements.

### Vegetation Protocol

The Vegetation Protocol collects standardized information on tropical forest trees and lianas. A TEAM Site will have between 6-9 1ha plots where trees and lianas larger than a pre-specified size are mapped, identified and measured. This results in each TEAM Site repeatedly measuring between 3000-5000 trees annually.

### Terrestrial Vertebrate Protocol

The Terrestrial Vertebrate Protocol collects standardized information on mid-sized tropical forest fauna (i.e. birds and mammals). This information is collected via camera traps (i.e. digital cameras with motion sensors housed in weather proof casings). The images taken by the camera trap are reviewed to identify what species are captured in the image by the camera trap. The image and the interpretation of what is in the image are the data for the Terrestrial Vertebrate Protocol.

The amount of data collected through the TEAM protocols provides a significant yet exciting IT challenge. The TEAM Network is currently partnering with the San Diego Super Computer Center to build the data management infrastructure. Data collected from the three core protocols as well as others are currently made available through the TEAM Network portal, which provides the content management framework, the data scheduling and management framework, an administrative framework to implement and manage TEAM sites, collaborative tools and a number of tools and applications utilizing Google Map and Google Earth products.

A critical element of the TEAM Network data management infrastructure is to make the data publicly available in as close to real-time as possible (the TEAM Network Data Use Policy: <http://www.teamnetwork.org/en/data/policy>). This requires two essential tasks to be accomplished, 1) A data collection schedule has to be planned, proposed and approved for a given TEAM site. This is a challenging process since TEAM sites are geographically distributed across the tropics and hence have different seasons where they schedule field sampling for the different TEAM protocols. Capturing this information and ensuring that TEAM sites follow the outlined legal contract is key to the data collection process and 2) A stream-lined and efficient information management system to ensure data collected from the field meet the minimum data standards

(i.e. are of the highest scientific quality) and are securely transferred, archived, processed and be rapidly made publicly available, as a finished consumable product via the TEAM Network portal.

The TEAM Network is achieving these goals by implementing an end-to-end framework consisting of the Sampling Scheduler application and the Data Management Framework.

#### Sampling Scheduler

The Sampling Scheduler is a project management, calendar based portal application that will allow scientists at a TEAM site to schedule field sampling for each of the TEAM protocols implemented at that site. The sampling scheduler addresses the specific requirements established in the TEAM protocols with the logistical scheduling needs of each TEAM Site. For example, each TEAM protocol defines when data must be collected (e.g. time of day, number of times per year, during which seasons, etc) as well as where data must be collected (from which sampling units, which trees, etc). Each TEAM Site has a limited number of resources and must create plans that will both satisfy the requirements of the protocols as well as be logistically feasible for their TEAM Site. With 15 TEAM Sites (and many more coming soon) the schedules of each TEAM Site must be communicated to the Network Office to ensure data are being collected as scheduled and to address the many problems when working in difficult environments like Tropical Forests. The Sampling Schedule provides built-in proposal and approval functionality to ensure that the TEAM Sites are and the Network office are in sync as well as provides the capability to modify schedules when needed.

#### The Data Management Framework

The Data Management framework is a three-tier data ingestion, edit and review application for protocols defined in the TEAM network. The data ingestion framework provides online web forms for field personnel to submit and edit data collected at TEAM Sites. These web forms will be accessible from the TEAM content management site. Once the data is securely uploaded, cured, processed and approved, it will be made publicly available for consumption by the scientific community. The Data Management framework, when combined with the Sampling Scheduler provides a closed loop Data Scheduling and Management infrastructure. All information starting from data collection plan, tools to input, modify and curate data, review and run QA/QC tests, as well as verify data are collected as planned are included.

Finally, TEAM Network data are available for download via the Data Query and Download Application. This application utilizes a Google Maps custom interface to search, visualize, and download TEAM Network data.

#### References

- TEAM Network, <http://www.teamnetwork.org>
- Center for Applied Biodiversity Science, Conservation International. <http://science.conservation.org/portal/server.pt>
- TEAM Data Query and Download Application, <http://www.teamnetwork.org/en/data/query>