A cyberinfrastructure for integrating data from an eddy covariance tower, robotic tram system for measuring hyperspectral reflectance, and a network of phenostations and phenocams at a Chihuahuan Desert research site

CYBER-ShARE Center of Excellence
Sharing Resources to Advance Research and Education through Cyberinfrastructure
THE UNIVERSITY OF TEXAS AT EL PASO

Christine Laney[1], Ryan Cody[1], Irbis Gallegos[1], John Gamon[2], Aida Gandara[1], Ann Gates[1], Libia Gonzalez[1], Jose Herrera[1], Aline Jaimes[1], Ari Kassan[1], Vladik Kreinovich[1], Ondrej Nebesky[1], Paulo Pinheiro da Silva[1], Gesuri Ramirez[1], Leonardo Salayandia[1], Craig Tweedie[1]. 1. University of Texas at El Paso, TX, USA 2. University of Alberta, Canada. Please contact cmlaney@miners.utep.edu or ctweedie@utep.edu for more information.

Introduction

The Systems Ecology Lab, with the assistance of the Cyber-ShARE Center (both of which are at the University of Texas, El Paso) is studying land-atmosphere carbon, water, and energy fluxes at a mixed creosote (*Larrea tridentata*) - mesquite (*Prosopis glandulosa*) shrubland in the northern Chihuahuan Desert on the USDA ARS Jornada Experimental Range. The site includes an eddy covariance tower built to Ameriflux and FLUXNET specifications, a robotic cart mounted on a 110m fixed rail system that measures hyperspectral reflectance, an 8-node network of SpecNet phenostations that match national phenology network standards. The site also includes other sensors, phenology cameras, and transects where the phenology of key plant species are monitored.

After more than a year of data collection, the lab has started to analyze the annual dynamics of reflectance and phenology from each of these data sources and to compare data between the sources. Each data stream is complex, requiring a number of processing steps and specialized software, including programs developed in-house by Cyber-ShARE computer science students. Integrating and analyzing the data to form a complete information system for measurement and synthesis of land-atmosphere carbon, water, and energy fluxes at the site provides a greater challenge overlying that of understanding data from each separate stream.

Currently, researchers (i.e. five graduate students) mostly follow a traditional mode of data management: each researcher manages his/her own datasets independently and manually produces processed data products as need dictates. In order to make data more easily accessible for processing, visualization, and sharing, much of this work can be automated. Here, we present a novel local cyberinfrastructure (CI) to help document, manage, visualize, and integrate high temporal and spatial resolution data from the multiple sensing platforms at the site. This poster provides an overview of this CI and is coupled with live demonstrations of the various tools that comprise it.

Research Site

The Systems Ecology Lab's Chihuahuan Desert research site is located on the USDA ARS Jornada Experimental Range, north of Las Cruces, NM (Fig. 1). It is a relatively species-poor site, dominated by creosote (*Larrea tridentata*) and mesquite (*Prosopis glandulosa*), with some tarbush (*Flourensia cernua*), bush muhly grass (*Muhlenbergia porteri*), and fluffgrass (*Dasylochloa puchella*). Soils are sandy and gravelly. On average, the Jornada receives 250 mm of rain annually, mostly during summer monsoon events. Rain that is not immediately absorbed by the top soil runs off of the site, eventually into arroyos. The site gently slopes predominately down to the west. Between the tower, tram, and sensor network, there are >100 sensors in operation, most continuously. Data streams, including 10hz eddy covariance data, are sent back to the SEL at least daily via wireless internet.

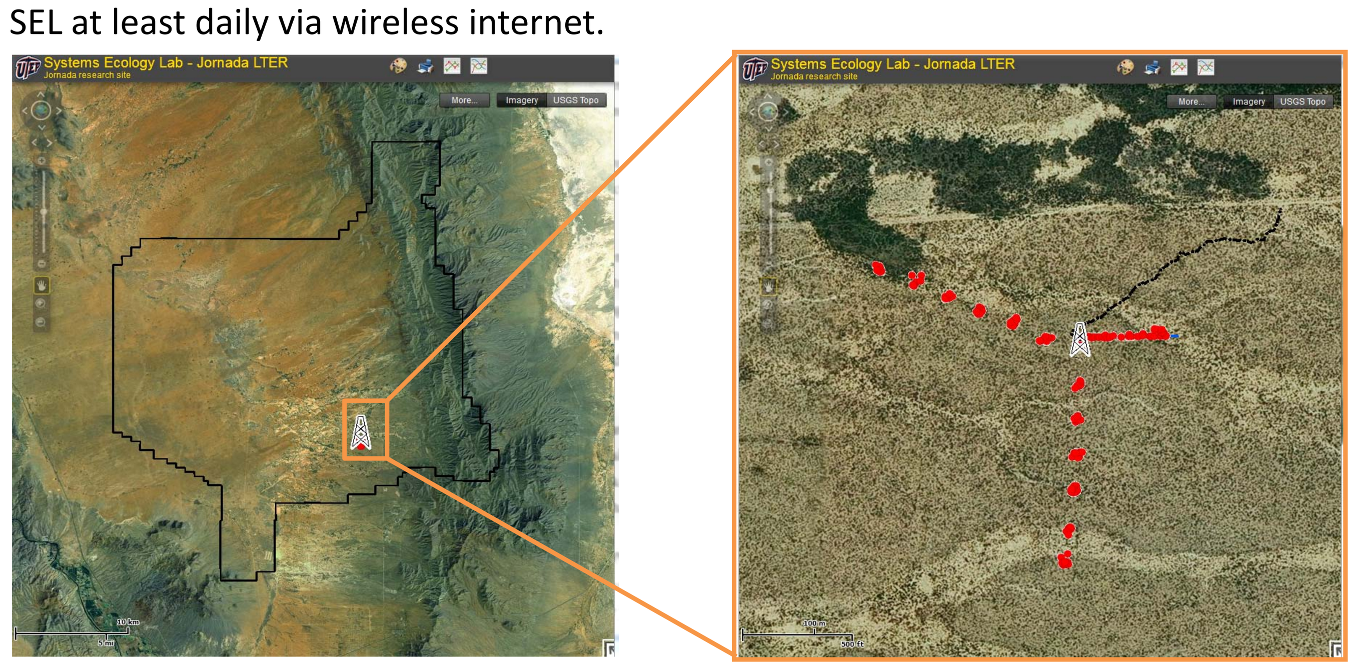


Fig. 1. Left: The Jornada Experimental Range, framed in black outline. Right: The SEL research site. Phenology transects are shown by red dots, the tram rail by a blue line, and the trail to the site by black dots. The tower is in the middle.

Research Questions

Some research questions posed by this site are:

1. What are the gross and net ecosystem production (GEP & NEP) of this system? (Please see Aline Jaimes' poster).
2. How is carbon, water and energy cycling at this site influenced by weather patterns and events?
3. What is the minimum set of sensors needed to estimate carbon, water and energy cycling in this system?
4. What is the optimum cyberinfrastructure needed to locally manage multiple complex data streams over time in a small, rapidly changing lab, and to share these data with multiple network partners?

Vision of the Cyberinfrastructure

A cyberinfrastructure (CI) is composed of multiple networked computers, software, information, and people. With the multiple large, complex data streams from the >100 sensors at the Jornada research site, sometimes collected at a rate of 10 values per second, it is unreasonable for SEL researchers to process and analyze the data without programmatic help. The vision of the future CI, currently under development, is a system that will:

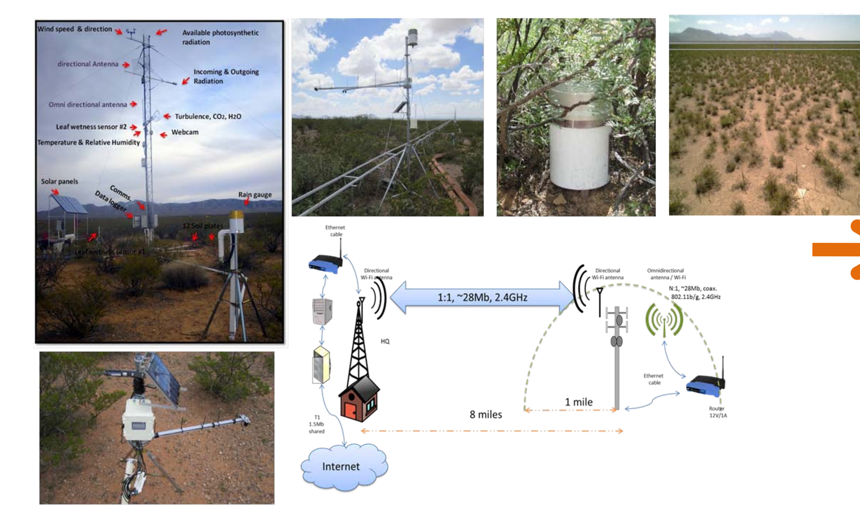
1. ***Periodically check each data stream for errors and flag all errors that are found.*** This would include a system to alert researchers to such errors, particularly those that indicate a failing sensor.
2. ***Store data in a consistent and reliable way***, such that researchers are not creating their own file names and manually manipulating raw data files.
3. ***Process data to generate derived datasets*** that can be used for analysis, by automating scientific workflows that may use one or more software packages, such as R or MatLab.
4. ***Insert new derived data on a periodic basis into a PostgreSQL database***
5. ***Document the data*** via standard metadata formats like OpenGIS® Sensor Model Language, as well as machine-accessible semantic web formats like Provenance Markup Language (PML).
6. ***Document the senor network and related field hardware components architecture***. Keep track of sensor installation, calibration and maintenance of system infrastructure.
7. ***Document scientific workflows*** for the benefit of current and future researchers, as well as collaborators who wish to use the data.
8. ***Visualize both historical and near-real time data*** within images of the spatial context that the data were collected.
9. ***Make the data available online*** for queries, analysis, download and visualization via web tools.
10. ***Periodically send data to other relevant data stores***, such as the FLUXNET and US National Phenology Network (NPN) databases.

Current Components of the Cyberinfrastructure

Several in-house software tools for information management have been tested and modified using tower data, some are currently in development, and others, particularly web services, are under discussion. Below are outlined the main components of the CI to date.

1. ***Common MS Access and PostgreSQL databases***: Researchers currently analyze and summarize their data using desktop applications such as MatLab or Excel. They load their summarized data into an Access database. These finalized tables and views are imported into a PostgreSQL database where they can be accessed for graphing and mapping.
2. ***Common basic metadata***: Each researcher has filled out a basic ASCII readme file with common fields describing data tables and how final data products were derived.
3. ***MatLab routines that estimate productivity from phenocam imagery*** (L. Gonzalez & Geovany Ramirez).
4. ***Data quality checking and flagging tools and machine learning algorithms***. Two software tools have been created to quality check tower data. The **Da**ta **Pro**perty Characterization Tool (*DaProS*), and the **S**ensor **D**ata **Ve**rification (*SDVe*), developed by I. Gallegos, are used in conjunction to specify data properties and verify data quality according to those properties, respectively. (Please see Aline Jaimes' poster for more information on these tools.)
5. ***Semantic web tools for documenting and sharing information provenance:*** Semantic Abstract Workflows (SAWs) are captured using the Cyber-ShARE developed WDo-IT tool (developed by L. Salayandia), and PROBE-It is used to compile the information into Provenance Mark-up Language (PML; P. Pinhiero da Silva). These combined efforts will eventually make data available for automated discovery and analysis on the world-wide web. (Please see Aline Jaimes' poster.)
6. ***An Adobe Flex framework for web based mapping and data visualization.*** The Flex framework uses spatial base layers from ESRI, unmanned aerial vehicle (UAV) images from the JER, and custom layers from the SEL. Flex widgets, customized by SEL, provide dynamic graphing and photo viewing capabilities (R. Cody, A. Kassin).
7. ***A custom data-handler*** that provides database querying, and web-service generation for interfacing with other CIs. This tool, called Joggler (O. Nebesky), uploads CSV and XML data files, caches the data in a PostgreSQL database, and makes these data available for querying by web services.

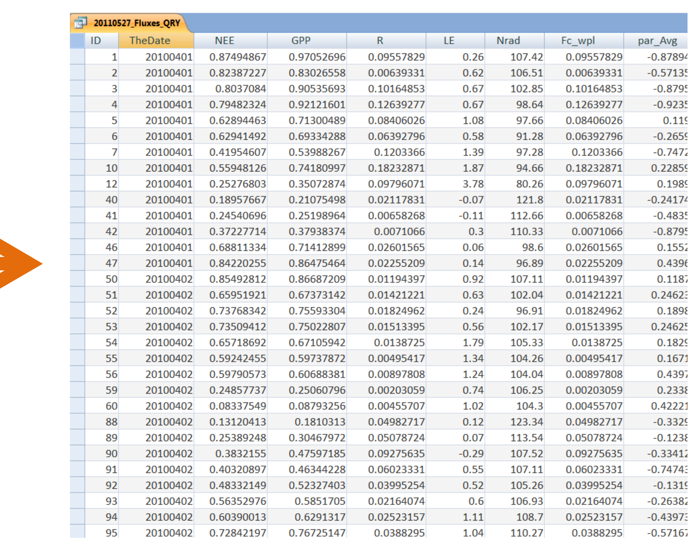
The CI – now and in the future



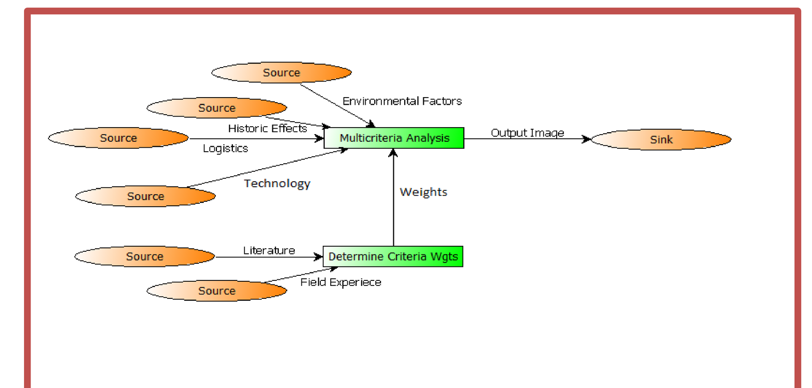
1. Tower, tram, phenocams, and sensor network send data wirelessly to a remote server.

The black box of data processing: MatLab, Excel, R, MultiSpec, etc.

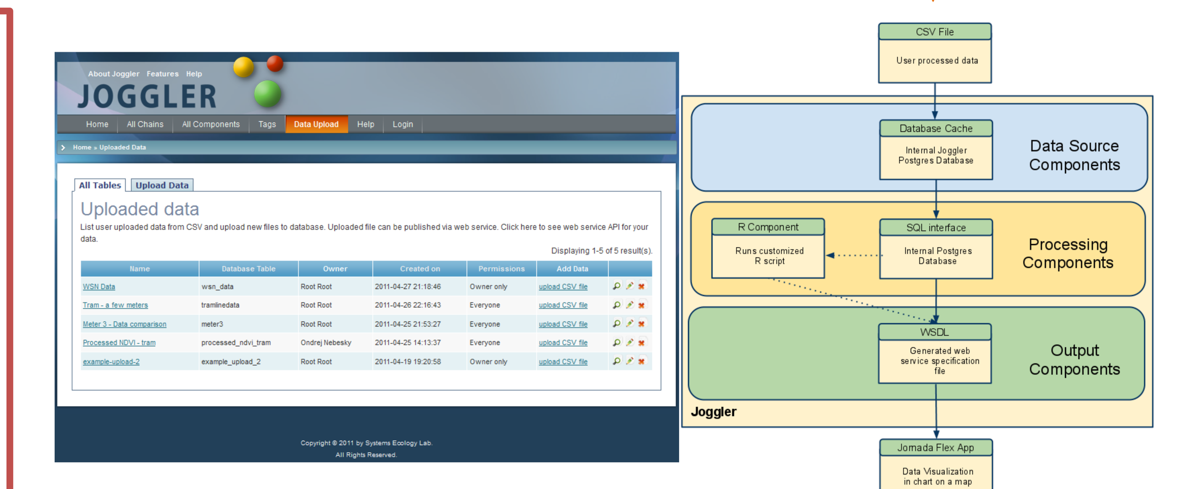
2. Data processing is currently manual via a variety of programs; to be streamlined with automated processes.



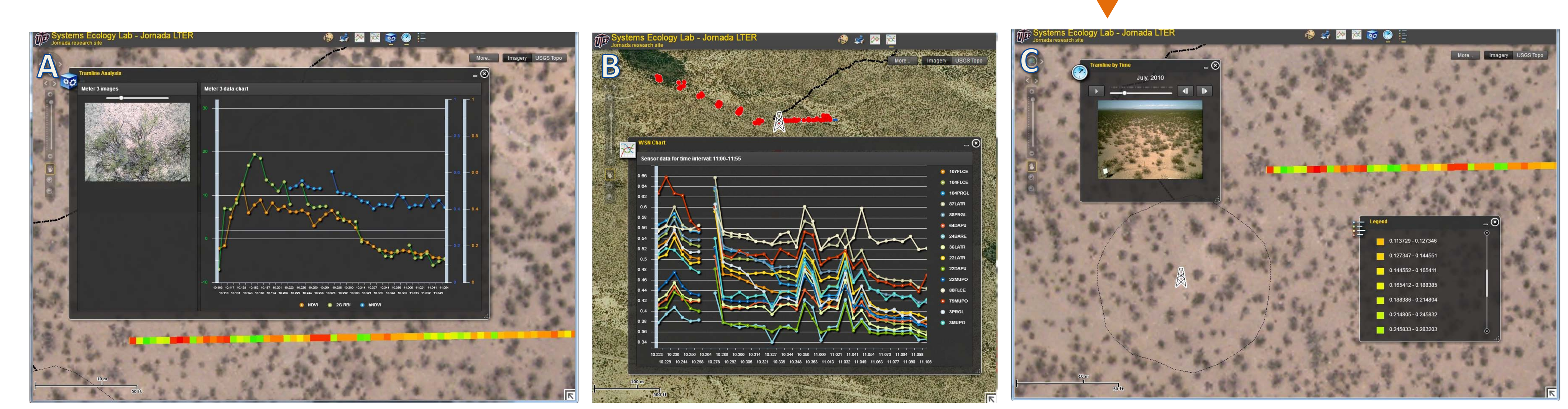
3. Researchers align their manually processed derived data and import it into an Access database



Some workflows and data products are currently documented as Semantic Abstract Workflows (SAWs) with WDo-It, but these are not as yet incorporated into the CI.



4. Joggler is used to import the data into a PostgreSQL database, where it will be available for the Flex website via web services.



5. Data are visualized via a Flex website with ESRI base layers, UAV high-resolution aerial photography, and SEL layers and temporal data (see Fig. 1 for symbology information). From left to right: A) Comparison of 3 comparable indices of productivity at one point along the tram line, calculated from reflectance sensors mounted on the cart (NDVI), on a sensor network node (bNDVI), and a phenocam (greenness). Photo on the left is from the phenocam mounted on the cart, corresponding to the meter mark at which the data were collected. B) Comparison of bNDVI from all solar radiation sensors mounted on the sensor network nodes. C) A time slider showing changes in productivity at each meter along the tram over time.

Future Work

Many of the visualization tools are currently in development as modifications to existing Flex or Flash widgets. However, much work needs to be carried out to document the data and scientific workflows for the semantic web, and to automate many of the data processing steps. Programming will also be needed to automate the push of data to various data partners such as FLUXNET and the US NPN. We are moving towards registering this site with national and international networks now that we have a year of data. We are very open to sharing data, CI tools, and other project products with other investigators and welcome opportunities for collaboration and constructive feedback. This project will be the focus of C. Laney's PhD research.

Acknowledgements

This research project, through Cyber-ShARE, is supported by the National Science Foundation under Grant No. HRD-073482. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The Jornada Basin LTER and the USDA ARS Jornada Experimental Range provided the research location and support. Fred Huemmrich and Steven Roach provided expert advise on data integration and analysis. Geovany Ramirez supplied MatLab code to analyze phenocam images.