The Canopy Database Project: An Informatics Resource for LTER Information Managers and Researchers

Organizer: Nalini Nadkarni and Judy Cushing

Key Participants: Nalini Nadkarni, Judy Cushing and LTER Information Managers: Kristin Vanderbilt, Nicole Kaplan, Eda Menendez, Ken Ramsey, Jonathan Walsh

September 19, 2003

INTRODUCTION

The management, analysis, and visualization of forest canopy data require specialized database tools that can be extended to broader ecological applications. In 1997, a group of forest canopy researchers and computer scientists created the Canopy Database Project to create informatics tools that would enhance the capacity of canopy researchers to manage their data at all stages of the research process. We have categorized canopy structure into a system with a finite number of "canopy space descriptors". This provides a framework to work at multiple spatial scales and facilitates cross-site comparison of forest structure. We have also developed a number of database design components, aka templates for efficient handling of complex data for data analysis and visualization. In May 2003, as a proof-of-concept activity, a two-day workshop with five LTER Information Managers, suggested novel cross-site comparisons and syntheses.

PARTICIPANTS

A total of 21 individuals (including the key participants listed above) attended all or part of the workshop. The majority identified themselves as informatics people, but at least four senior ecologists participated. Attached is a list with contact information to facilitate future communication.

ACTIVITIES AND RESULTS

In this three-hour workshop, we described our database, which is comprised of forest structure/function studies from multiple Principal Investigators. We presented artifacts generated by our database tools, which include DataBank (a tool for database design) and CanopyView (a tool for data visualization). We discussed potential uses of these software tools by LTER researchers and Information Managers, and determined what needs to be done to make these more useful to the larger ecological community.

We presented the value of complementing software components (templates) with semantic information and explored ways of discovering documenting concepts and their component structures using processes that have been successful in other domains. A preliminary categorization of canopy

spatial structures and preliminary concept maps of primary productivity for grassland sites were presented. Presenters included Nadkarni, Cushing and Kaplan.

For part of the workshop, the full group broke into two smaller groups, one to focus on forest and canopy structure issues, and the other to discuss the informatics work associated with assessing net primary productivity using LTER databases. The two groups then reported out to the full group on their results and discussions.

The **forest canopy structure group** (5-7 individuals) discussed the value of working with such a system, and agreed that it was a useful approach. Several examples of forest space categories were generated from data collected at LUQ. Different types of coordinate spaces were discussed, noting the multiple ways that a single data set could be both conceptualized and displayed. The ability to "stack" other data sets on the original data was worked out. Visualization (or other representations of data) was viewed as a key part of this process, and as a powerful reward for using this system.

The cultural aspects of these issues were also discussed, i.e., how to either give rewards and incentives to scientists to deposit and share data, or to provide disincentives NOT to share data. One of the "carrots" that could be offered is the greater ease with which collaboration and synthesis could occur if the canopy database and the forest categories were used by individuals and by the research community.

The **grasslands net primary productivity group** (7 individuals) used the preliminary concept map and knowledge of data at three LTER sites (JRN, SEV, SGS) to develop preliminary domain models covering relevant data at those sites. The three domain models were later refined and we determined that they were likely comparable. From those domain models, Databank templates will be devised and an experiment in integrating these data into one synthetic data store performed.

NEXT STEPS AND FUTURE COLLABORATIONS

We have organized a follow up 2-day workshop that will be held at the LTER Network office in New Mexico during the third week of November, 2003. At this workshop the DataBank templates for grassland net productivity will be used to generate three databases, and those databases populated. We will then aim to generate one database from those that could be used for ecological synthesis.

We will also continue to interact with forest researchers in our efforts to categorize canopy structure. We have a full year of funding from our current National Science Foundation grant to refine our ideas and build a larger database to support these efforts. We anticipate that we will apply for a renewal grant (for two subsequent years of funding) to enhance our canopy database project, particularly aspects of canopy structure conceptualization, visualization and statistical analysis.

We also anticipate working closely with SEEK to - cast our canopy structure concepts into formal concept maps, using tools they have identified as of potential value to ecology. Finally, we hope to work collaboratively with NCEAS to develop synthetic and integrative work that will grow from our canopy database activities.

CONCLUSIONS

The LTER All Scientists Meeting was a valuable venue to disseminate information about the canopy database project and forest canopy structure categories, and to seek feedback from ecologists not previously involved in the project. Both the project and the concept categorizations seemed to appeal to LTER scientists, as the topic provoked interest and useful feedback. The responses confirmed the PIs assessment that the use of software components to share, visualize, and analyze data will be a useful activity to individual scientists and to the forest ecology research community. Visualization in particular was deemed a most important draw. The prospect of carrying out synthetic studies more easily also makes participation in this database project compelling.

ACKNOWLEDGMENTS

PARTICIPANTS IN THE CANOPY DATABASE WORKSHOP			
September 19. 2003 Seattle, Washington			
First Name	Last Name	Site/Institution	email contact
Judy	Cushing	Evergreen State Coll.	judyc@evergreen.edu
Mike	Figgens	SEV	friggenj@sevilleta.edu
Mike	Finch	Evergreen State Coll.	fickerm@evergreen.edu
Julian	Hadley	Harvard Forest	jhadley@fas.harvard.edu
Clair	Hoff	U of New Hampshire	choff@cisunix.unh.edu
Nicole	Kaplan	Shortgrasse Steppe	nicole.kaplan@colostate.edu
Nalini	Nadkarni	Evergreen State Coll.	nadkarnn@evergreen.edu
Julius	Oszlanyi	Slovak Academy	juliu.oszlanyi@savba.sk
Tim	Schowalter	Louisiana State Univ	tschowalter@adcenter.lsu.edu
Paul	Schwarz	Oregon State Univ	paul.schwarz@oregonstate.edu
Jason	Walker	Central AZ	jasonwalker@asu.edu
Wu	Wu	LUQ	wwu01@syr.edu