NETWORKING URBAN ECOLOGICAL MODELS THROUGH DISTRIBUTED SERVICES

Peter McCartney[1], Ray Quay[2], Corinna Gries[1], Charles Redman[1], Tony Fountain[3]

[1]*Arizona State University;* [2]*City of Phoenix;* [5]*San Diego Supercomputer Center.*

http://ces.asu.edu/itr

Understanding of the functioning of urban ecosystems cannot be achieved without the data collection and modeling developed by scientists working for federal, state, and local agencies, often in cooperation with university researchers. The project engages a broad range of partners. Local government agencies include the City of Phoenix (COP), Maricopa Association of Governments (MAG) and Arizona Department of Water Resources (ADWR). These and other agencies have been engaged through the Greater Phoenix 2100 organization (http://gp2100.org), a multi -agency partnership led by ASU to identify long-term planning issues. Collaboration with the Central Arizona Phoenix Long-Term Ecological Research (CAPLTER) project links this effort with an energetic research program in urban ecology. Partners in informatics research include the San Diego Supercomputer Center, which participates in a subcontract, and collaborative interactions with the Science Environment for Ecological Knowledge (SEEK), and the Geosciences Network (GEON) in the form of shared software and metadata standards, joint meetings and working group memberships.

The Networking Ecological Models project seeks to 1) establish a multi-agency network of metadata, data, and application services that can be invoked through an open, platform-independent messaging format and 2) enable the creation and execution of scenarios, or workflows, that loosely couple models by "piping" outputs of one process to the input of another – even if the two processes are running in different locations, on different operation systems, or in different languages. To test our infrastructure, we are implementing a workflow that couples output from a global climate model to a groundwater model and to a local model of urban growth (Figure 1).

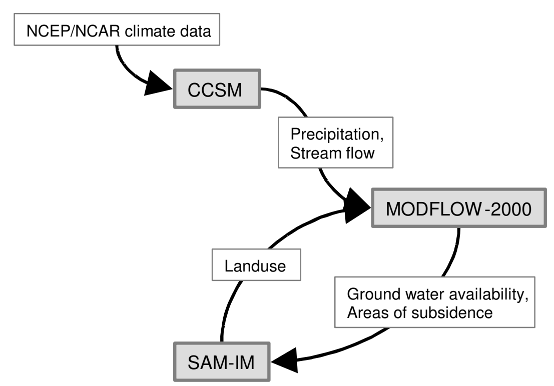

Figure 1. Data flow and coupling of evaluation models.

Our objectives for reaching these goals have been to leverage on existing infrastructure while at the same time incorporating technological solutions under active development by our collaborative partners. Thus, a core principle for our design was to focus on select models that are widely accepted and actively used in urban planning today. Under prior research, CES has developed a query system for distributed metadata catalogs and coupled this with data query services. The current project is extending these resources in several ways. While our project is very much a proof of concept, we have sought to implement certain elements that are clearly emerging as critical components of global cyberinfrastructure in the science, business, and government domains. These include web services, metadata standards, and distributed grid computing, and workflow processing.

Our accomplishments to date have been to define a basic interface design for our web service interfaces for data access and processing, establish an architecture for how these services interact with each other, and to implement them in a local, single-node environment using hand-crafted java classes to handle the logical flow. We are now replicating this architecture on multiple nodes, or points-of-presence (POP) and are implementing additional services to manage the transport of large datasets between each POP, and to create the web service interfaces for requesting model execution at those nodes. At the same time, we are working (through our SDSC partnership) to adapt workflow processing tools from the Kepler project, a large collaborative effort to extend the Ptolemy II workflow processing software to support grid and distributed computing functions.

The project management structure is designed to cross-cut institutional boundaries. The metadata design team is linked with Ecoinformatics.org, a partnership that maintains Ecological Metadata Language (EML). CES collaborates closely with MAG, ADWR and COP in the planning of the use-case test application as well is in the linking of the system to the agency datasets and models. Technical development on web service design and the workflow processing tools is carried out in close partnerships with SDSC, the ASU participants in the GEON project and members of the SEEK project.

The project has also focused on several short-term goals that have already succeeded in engendering inter-agency interest in information infrastructure for environmental data-sharing. One has been the continued extension of the Southwest Environmental Information Network (http://seinet.asu.edu) to add metadata catalogs of available datasets targeted as relevant to an inter-agency dialog on water resources and water planning moderated by ASU's Consortium for the Study of Rapidly Urbanizing Regions (http://ces.asu.edu/csrur). We also collaborated with the GP2100 project to develop the GP2100 eAtlas (http://www.gp2100.org), an internet map atlas that is directly coupled to distributed, online data archives and has proven effective in promoting public awareness of long term environmental that are affected by policy decisions. Our efforts in model metadata definition have also led to new connections including a joint workshop in August 2003 with a partnership dedicated to forming an open-research community in ecological modeling following the principles of open software development. A final workshop will be held this year to finalize a draft proposal of the model metadata to EML.

The broader impacts of the project have been to enhance the dialog between ASU and government partners on goals for data sharing and integration. Since its award, numerous new grant proposals have been generated that build upon the infrastructure goals of this project and which further engage its participating members. Through the eAtlas and the model integration goals, the project has contributed to the definition of several major thrusts for ASU including a Decision Theater project for visualization and decision support, and the Consortium for the Study of Rapidly Urbanizing Regions; and for SDSC such as the development of the Computational Urban Environments initiative.

The project has faced a number of challenges. Close working relationships with agency IT staff and with remote partners is essential to forming the semantic connections between academic research and government planning goals. We have found it is very easy to let communications drop and would build stronger mechanisms in the future to make them more robust in the face of inevitable scheduling, personnel change, and strategy shifts. We also have dealt with the issues of balancing between completing prototypes versus trying to track the pace of new developments in cyber infrastructure. What has emerged as a clear signal from this project is that major environmental problems are not “owned” by single agencies. Long-term decisions require integration of information in a manner that leverages, rather than competes with the infrastructure, models and metrics that are currently used by decision-making groups