Integrating Urban Ecological Models

Peter McCartney[1], Ray Quay[2], Corinna Gries[1], Amy Sundermier[1], Robin Schoeninger[1], Joseph Zehnder[1], Anubhav Bagley[3], Dale Mason[4], Longjiang Ding[5]

[1]Arizona State University; [2]City of Phoenix; [3]Maricopa Association of Governments; [4]Arizona Department of Water Resources; [5]San Diego Supercomputer Center.

Over the past 20 years, understanding of the functioning of urban ecosystems has increased to the point where fairly sophisticated models of various subsystems have been developed by scientists working for federal, state, and local agencies, often in cooperation with university researchers. Each is designed to collect data and explore a specific urban/environmental system for issues related to the mandate of the sponsoring agency. These models are of value to science both for their insight into social and economic process as well as potential media for ecological monitoring data to be represented in environmental planning and decision-making.

One challenge facing us is the development of an information infrastructure that will facilitate the sharing of models among the diverse members of the urban ecological research community in central Arizona. In this project we build on prior work that established the Southwest Environmental Information Network 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.

As a test of this system, we have defined a workflow that links output from a global climate model (CCSM2) running at National Center for Atmospheric Research to a hydrology model (MODFLOW2000) running at the Arizona Department of Water Resources and then to a land-use change model (SAM-IM) developed by the Maricopa Association of Governments as an ArcView GIS model (Figure 1). Coupling these models in an integrated workflow involves accessing source data in several formats, executing models in different locations, capturing outputs and transforming them to overcome scalar and semantic differences between the outputs of one model and inputs to another, and moving the result to the appropriate location to be used as input to the next analytic process. A workflow-processing engine interprets the script and generates the specific calls to retrieve data, transform data, and execute models.

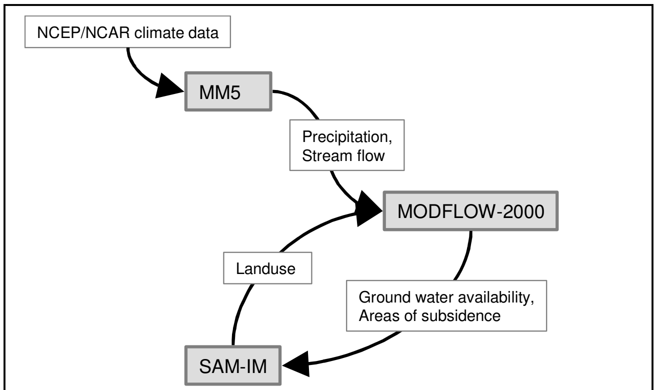

Figure 1. Data flow and coupling of evaluation models.

This poster illustrates the underlying architecture of the integration network in which each agency participates as a node within a local grid environment by hosting a compatible set of web services to provide metadata search, data query and processing, and application access (Figure 2). A standardized message structure creates a uniform interface to all application components. Data provided by an agency are documented in a local metadata catalog using Ecological Metadata Language (EML). Each processing service is designed to read EML to read incoming data and generate EML to describe its output. In most cases, the data processing services are actually wrappers to existing software such as the

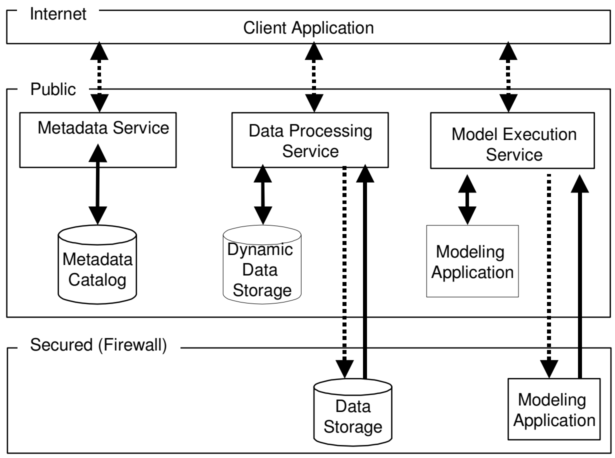

Figure 2. Software and data components of a network node. Solid lines represent restricted communication; dashed lines represent SOAP messaging.

GRASS GIS package or R Statistical package. The same approach is taken to wrap the models themselves. The wrappers serve to support a common message format for scripting workflows and executing them from a single process. In developing the prototype for this project, we are making use of workflow processing software based on the Ptolemy II application. Several ITR-funded projects, including this one, are collaborating to extensions to Ptolemy for integration with web and grid service components, as well as a broad range of ecological modeling and data functions. In this project, we are developing support for distributed workflows though FTP transfers of data between node drop-boxes on the network and structured metadata to pass relative addresses of data to the remote processing service. By distributing the computing software, it is possible to generate workflow scripts that take advantage of the co-location of data and processing code, minimizing the size and frequency of data transfers between two different nodes.

Two important issues affecting an agencies participation in an open data and application sharing system are 1) security and 2) increased load due to external access traffic. To mitigate these two issues during prototype network development, servers were acquired by the project and configured with metadata catalogs and web service components. These were then installed at the two participating government agencies as independent servers from their normal computing resources. Placement of these systems outside firewalls eases limitations on network access and allows agencies to control the communication between the node server and their internal resources.

The significance of this approach to integration is that it leverages existing investments in models and data management systems. We expect this prototype to lead to an infrastructure that will support more rapid and flexible collaborations between academic and government agencies by reducing the need to move copies of large datasets to new locations or recode existing models to comparable languages.