Data Plan

All laboratory results, field observations, simulation results and the input data used for this proposed study will be made available after the study is complete. These lab, field and model inputs and outputs will be kept electronically in a repository in such a manner to both archive the results and provide them to the scientific community. Owing to the great volume of data created by this project and the desire to make that data available once the project is complete, a database will be set up, complete with some simple tools to navigate it and a compute server dedicated to that purpose following the template of Program for Climate Model Diagnostics and Intercomparison which archives a large number of global climate simulations used in the IPCC Assessment and provides the scientific community access to them via a website and data portal. The availability of this web portal will be mentioned in publications that result from this work and PIs, students and postdocs will endeavor to make simulation results transparent and of use to researchers, regulators, managers and the public.

Web portal

A website will be setup and maintained for this project that will provide a project summary and links to the collections of input data and model output. Because vast amounts of data will be generated through observations and model simulations, if it is not possible to store all the data on the web server, a mechanism will be provided to request collections of data through email or a web form. The PIs will extract the appropriate data from the file server (discussed below) and forward to the interested parties. External links will be provided to fully document the input data sources. The model output archive will cross-reference publications and associated documentation to provide a full description of available data.

Data storage

We anticipate that the online and offline coupled model simulations will generate terabytes of data that will need to be archived for further processing and data analysis. This data will be stored on a parallel file server most likely hosted by NCAR through their new High Performance Storage System (HPSS). Any simulations performed will be archived locally on this server to maintain a central database of simulation results. Conventions for directory naming and file structure will be established at the beginning of the project. The simulation codes used in the proposal have input and output files that are written in various high-level scripting languages. These languages will allow the scripts to categorize and archive both the model input and output (see below) automatically and provide automatic updates for the web portal and web log (see below). The web portal will contain lists of the available runs and directories with pointers stored to indicate the location of the data on the file server.

Input data

The coupled models used here require a vast array of input data sources. The web portal will describe the sources and data types and link to the appropriate directories on the file server. Web links will be provided to the original sources of electronic data so that users can obtain data from the same sources but for applications to other geographic locations.

Model output

All model output will be archived according to the model run names and linked from the web portal. Raw model output will include full three dimensional fields of the subsurface and atmospheric variables, together with stream flow data and other surface hydrologic and energy variables. These will be made available to facilitate new post-processing and analysis by interested researchers. In addition, our analysis techniques and processed model output will be made available. Codes and post-processing scripts written for data analysis will also be provided

with the processed data. Appropriate documentation for the post-processing codes will be provided.

Model code
The coupled codes are already publicly available. The precise versions of the codes used in these simulations will be archived such that they can made available for anyone to replicate project simulations. Version control for the code and post-processing codes will be managed using SVN (Subversion, the predecessor to CVS, Concurrent Versions System, open-source version control software: http://en.wikipedia.org/wiki/Apache_Subversion). The web portal will include an online viewer for the SVN repository for project users (and other collaborators as desired). Comments in the repository will be supplemented by discussions in the web log (described below). In addition to providing the code and post-processing scripts, input files for performing the production runs will be included. Input files for running simple verification tests will also be included. Relevant components of the codes used and developed under this project will also be transferred to the WRF and Noah model support groups at NCAR for further inclusion in community releases of these models.

File formats
Standard file formats (e.g. GRIB, NetCDF, ASCII, SILO, HDF) will be used whenever possible to make the files accessible from a range of platforms and applications. Pre- and post-processing software will also be made available when file formats are non-standard.

Project web log
A web log will be set up to document code changes, model runs, and model input/output details as they are developed. This will allow project participants (graduate students and supervisors and collaborators) to efficiently communicate and accurately record progress with simulations and changes to the code. This is essential for management of a large code (coupled groundwater and atmospheric components) and the large number of runs that will be performed. A simple web log is already available for ParFlow: http://parflow.blogspot.com/2007/09/welcome-to-parflow-blog.html. An enhanced version will be created specifically for this project.

Documentation and publications
A users' guide to the model code and the data will accompany each set of code and data (both input and output) and the model. This will be augmented by the detailed logs in the project web log and in the SVN repository. Model results and analysis will be disseminated widely through journal publications and conference presentations. Our results and data collected and produced will therefore be subject to peer review. We have requested funding to attend conferences that span disciplines.