Integrating Data, Models, Uncertainty Analysis Methods, and Super Computing to Facilitate Modern Environmental Assessments

Presented by Gerry Laniak
Ecosystems Research Division
National Exposure Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
October 12, 2009

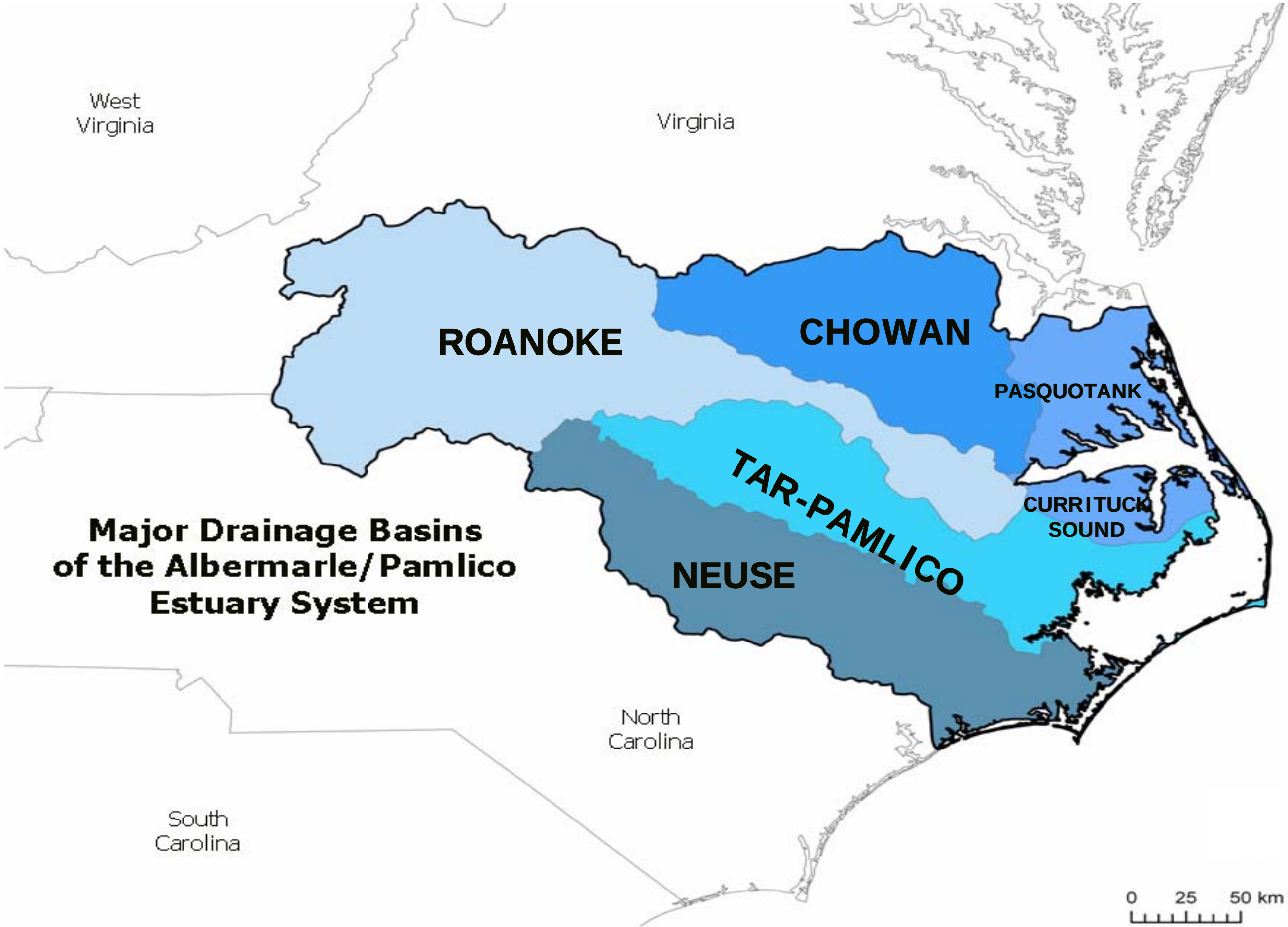
Presentation Goals

- Describe an example of a complex multi-media modeling problem
- Describe the software-based technology system designed to support such modeling

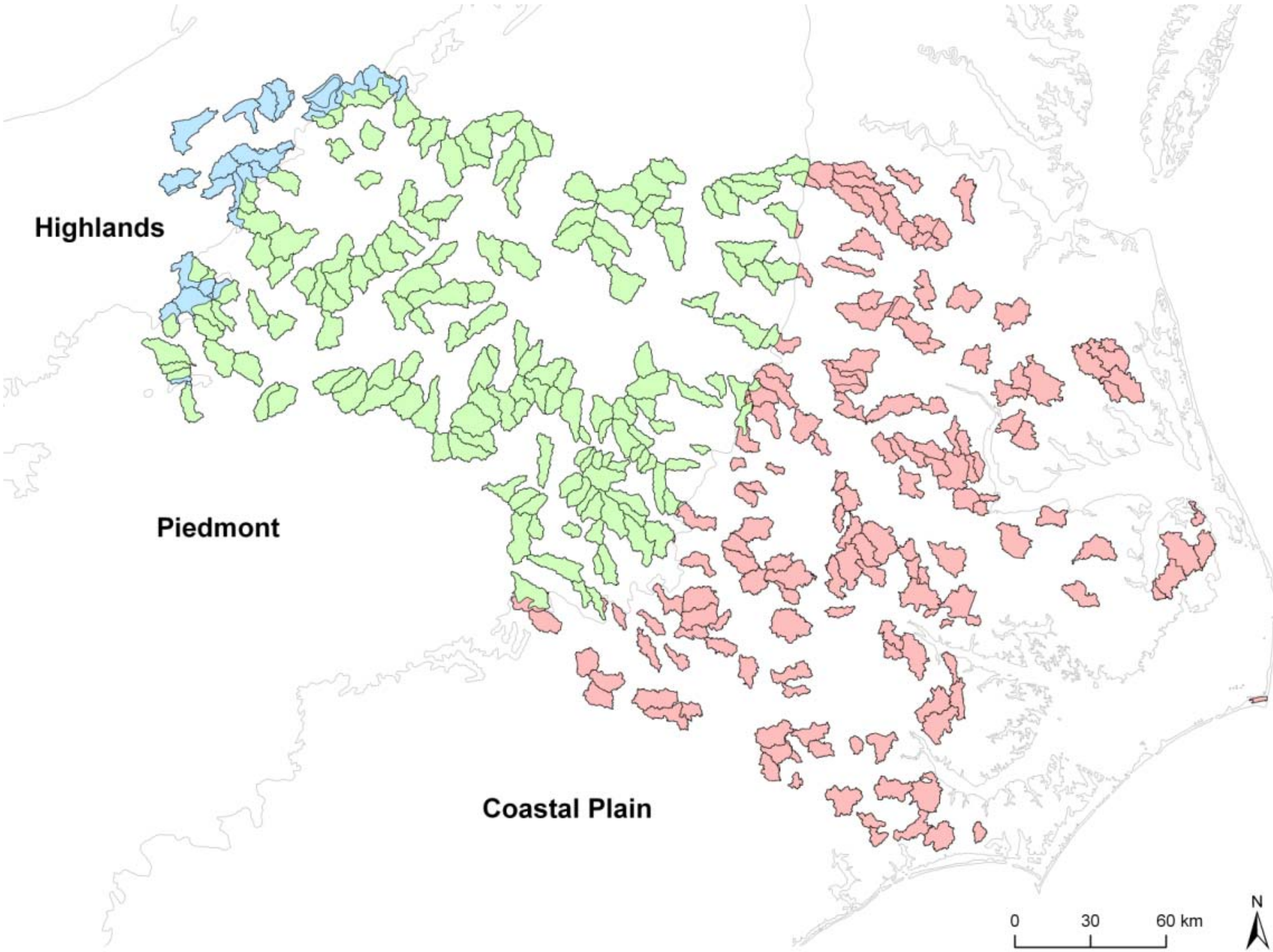
Conceptual Problem Statement

How will aquatic ecosystems and services related to fresh water recreational fisheries across a sub-regional to regional landscape be affected by changes in nitrogen, mercury, and pesticide loading patterns under various land-use and climate change scenarios?

Region of Initial Interest : Albemarle-Pamlico Estuary System (APES)



Definition and Specification of Fresh Water Fishery



Stressors of Interest
(Represented for Baseline and Alternative Future Scenarios

- Land use distribution across region
- Climate as represented by distribution of precipitation and water temperature throughout region
- Regional distribution of N loadings from atmosphere
- Regional distribution of Hg loadings from atmosphere
- Distribution of pesticide application rates across agricultural areas
- Distribution of N loadings from agricultural areas (crop/animal operations)
- Regional distribution of N loadings from point sources (treatment/septic systems)

Ecosystem Response and Service Measures/Indicators of Interest

- Water Quantity (flow, depth/velocity)
- Water Quality (water column and sediments)
  - DO, Chlorophyll a (water column only)
  - TOC
  - Nutrients (N,P)
  - Contaminants (Hg, Pesticide)
  - Temperature
  - TSS
- Aquatic Biomass and Productivity
  - Primary production (phytoplankton)
  - Secondary production (invertebrates)
  - Tertiary production (fish biomass/production for dominant/game/indicator species)
- Hg, Pesticide concentrations in fish
- Habitat suitability (as a function of productivity endpoints)

Two Types of Decision Level Assessment Questions We Want to Answer

(With Quantified Sensitivity & Uncertainty)

- **Type I.** What percent of fresh water fisheries in the APES are expected to demonstrate at least an *X* percent change in their provisioning of ecosystem service *S* in conjunction with stressor scenario A over the next 5, 10, and 20 years?

- **Type II.** What percent of fresh water fisheries in the APES are expected to have their provisioning of ecosystem service *S* below the threshold value of σ in conjunction with stressor scenario A over the next 5, 10, and 20 years?

Assessment Question Applied to : **Fishery Production**

- What percent of fresh water fisheries in the APES are expected to experience a reduction in annual production of at least 30% in conjunction with stressor scenario A over the next 5, 10, and 20 years? **(“fishery” production-Type I)**

- What percent of fresh water fisheries containing in the APES are expected to have an annual production of less than *X* g/m2/yr in conjunction with stressor scenario A over the next 5, 10, and 20 years? **(“fishery” production-Type II)**

Assessment Question Applied to : **Water Quantity**

- What percent of fresh water fisheries in the APES are expected to decrease their mean annual streamflow by at least 30% in conjunction with stressor scenario A over the next 5, 10, and 20 years? **(water quantity-Type I)**

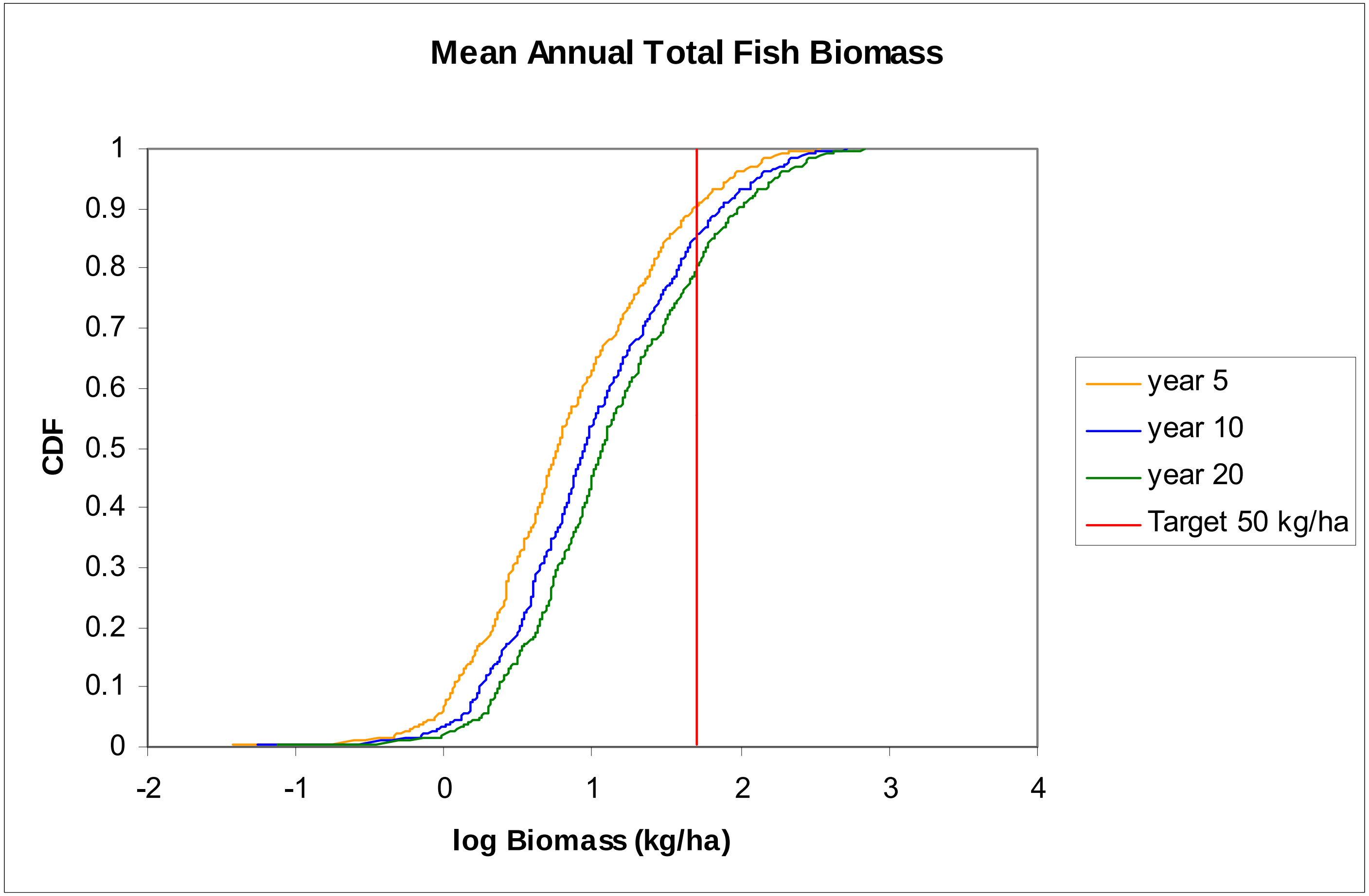
- What percent of fresh water fisheries in the APES are expected to decrease their mean annual streamflow to *X* m3/yr in conjunction with stressor scenario A over the next 5, 10, and 20 years? **(water qantity-Type II)**

Assessment Question Applied to : **Wildlife Habitat**

- What percent of fresh water fisheries in the APES are expected to have habitat suitability scores for game fish (or dominant/indicator fish) decrease by at least 30% in conjunction with stressor scenario A over the next 5, 10, and 20 years? **(wildlife habitat-Type I)**

- What percent of fresh water fisheries in the APES are expected to have habitat suitability scores for game fish (or dominant/indicator fish) less than 0.5 in conjunction with stressor scenario A over the next 5, 10, and 20 years? (**wildlife habitat-Type II)**

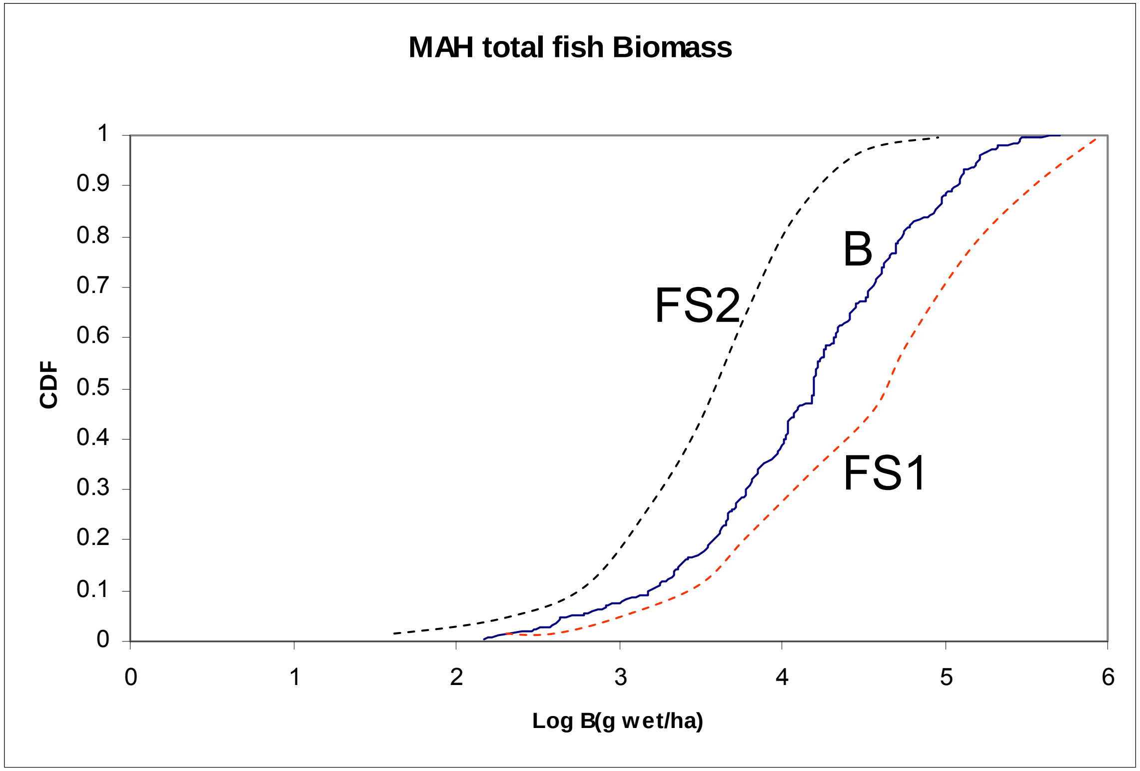
Illustrative Example of Regional Roll-up



Source: 25-yr BASS spin-up simulations of 363 EMAP Mid Atlantic Highland streams assuming repeated annual water temperature and discharge regimes.

Example Annual Fisheries Roll-up

Mean annual biomass, Year 5



OK, that’s the problem statement

Now a pathway to a solution

- Assessment Methodology
- Models
- Data
- Integrating Technologies

Essence of Assessment Methodology

- **Select and link existing and new models to form a modeling system to simulate HUC12 scale watersheds, connected surface waters, and fishery (pour point segment of surface water network).**

- **Select a statistical sample of headwater HUC12 watersheds for analysis**

- **Prepare datasets for each watershed and scenario of interest (baseline, alternative futures).**

- **Apply modeling system to each sampled HUC12 within a Monte Carlo simulation**

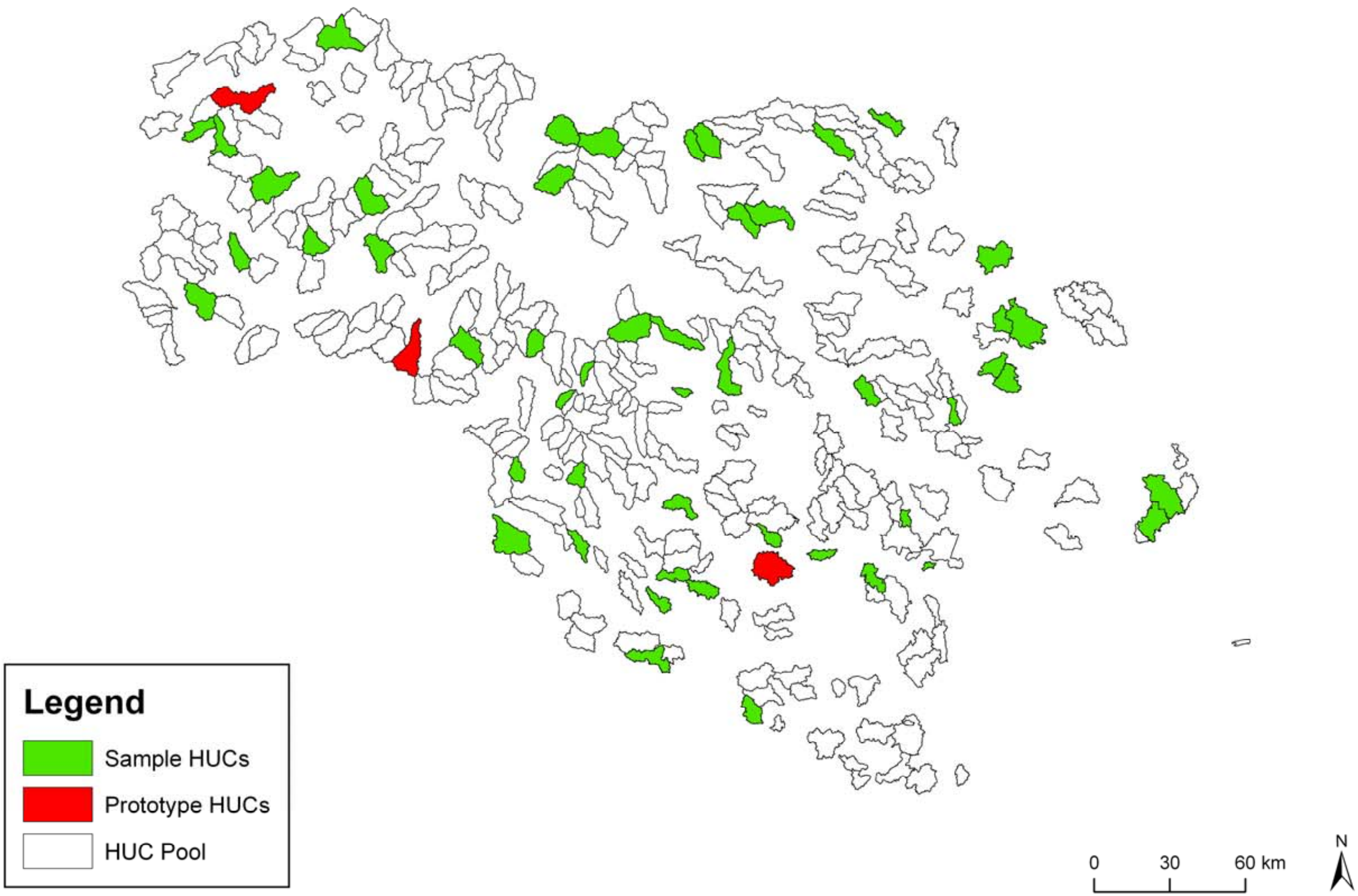
- **Process modeling results to calculate annual summaries.**

- **Collect results across HUC12s to develop regional distributions of water quantity/quality, habitat suitability, and biotic productivity and to characterize sensitivity and uncertainty.**

The Models

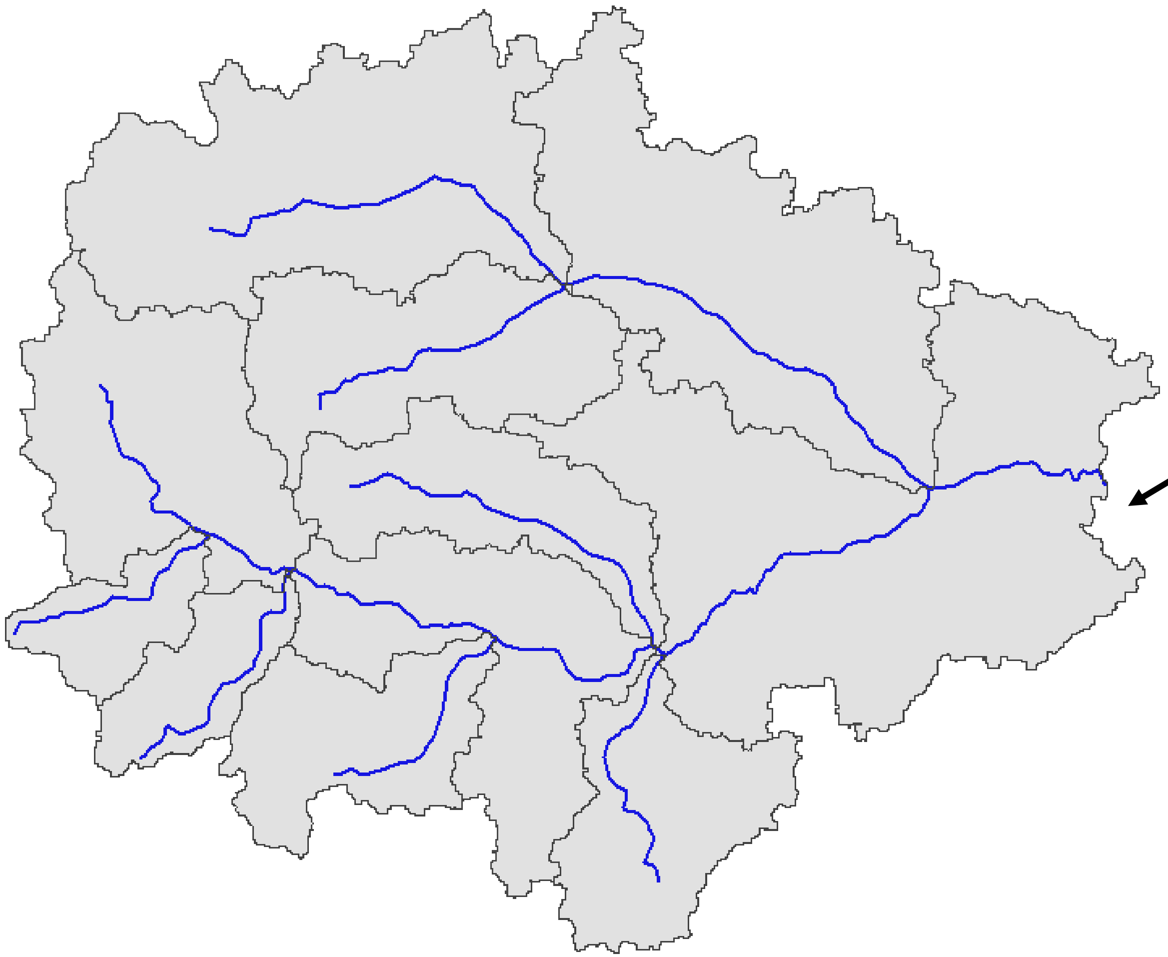
- CMAQ (Community Multi-scale Air Quality)
  - Regional Hg and Nitrogent deposition
- SWAT (Soil and Water Assessment Tool)
  - Watershed hydrology, sediment transport, agricultural processes, nutrient/pesticide fate and transport in the watershed
- Watershed Hg (new model for APES)
  - Hg fate and transport in the watershed
- WASP (Water Quality Analysis Simulation Program)
  - Water quality in the water column and sediments of stream network
- HSI (Habitat Suitability Index, new model for APES)
  - Habitat suitability for fish species
- BASS (Bioaccumulation and Aquatic System Simulator)
  - Fish population dynamices
- ESP (Ecosystem Service Processor, new for APES)
  - Modeling post processor
- UDP (Unit Definition Processor, data manager for APES)

Sample and Prototype Headwater HUCs



Unit of Analysis : 12-digit HUC

(Example for Middle Swamp, NC)



Freshwater fishery assumed at pour point of each headwater HUC12

Fish assemblage assigned to each Location based on regional cluster analysis

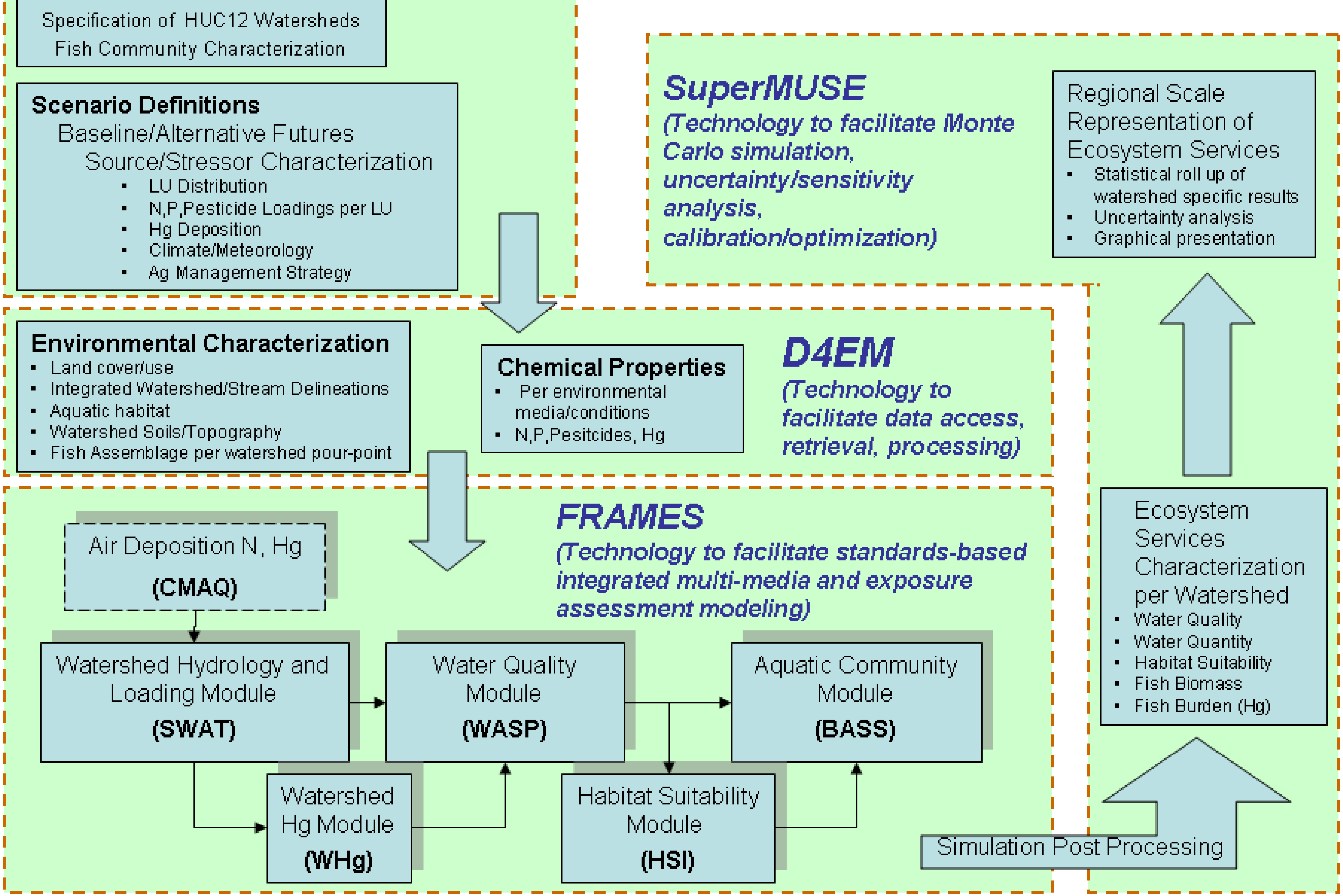
Data Needs for APES

- Data available from National data sources
  - –Meteorological data time series
  - –Watershed characterization
  - –Stream network
  - –Landuse/cover (including crops, animal operations)
  - –Soils data
  - –Chemical property data (Hg, Nutrients, Pesticides)

Data Needs for APES (cont'd)

- Data of a more site-specific nature
  - –Farming practices (e.g., tilling, fertilization, waste disposal)
  - –Fish communities and densities (1 community/HUC)
  - –Fish properties (78 species, 4 properties each)
  - –Background concentration load fluxes (66)
  - –Deposition data (2/HUC)
  - –Stochastic variable distribution parameters (89)

REGIONAL SCALE ASSESSMENT OF AQUATIC ECOSYSTEM SERVICES



Goals of Integrated Technology

- Design a comprehensive technological solution for APES that can be directly applied in other regions and whose software components can be reused for other problems
- Leverage/reuse existing open source technologies
- Automatically populate 100% of the data files
- Provide systems tools (e.g., MC, UA/SA, data viewers) that are directly applicable to other models and modeling systems
- Standardize the information flow through system
- Design for transparency, QA, reuse, and interoperabilty

Elements of Modern Environmental Modeling Systems

- Science-based models
- Large-scale environmental databases
- Assessment features (e.g., Monte Carlo simulation, calibration, optimization) *
- User interfaces *
- GIS-based data access, organization, viewing, and analysis *
- APIs for managing data within modeling system *
- Data analysis and visualization tools *
- Distributed Computing tools *

* framework/infrastructure, i.e., support software

Definitions

FRAMES : **(Framework for Risk Analysis for Multimedia Environmental Systems)**

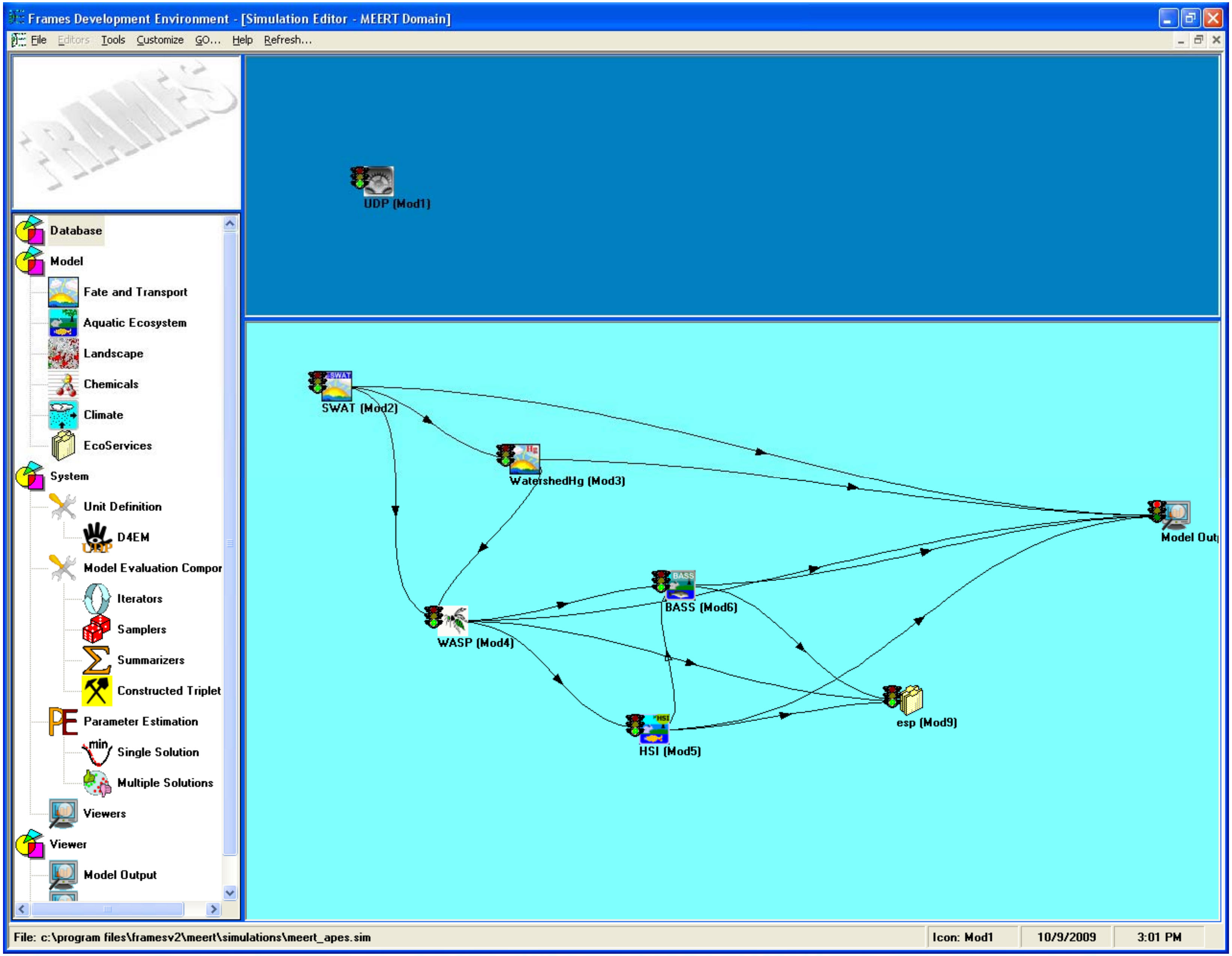
a software system that facilitates the linking and execution of individual models

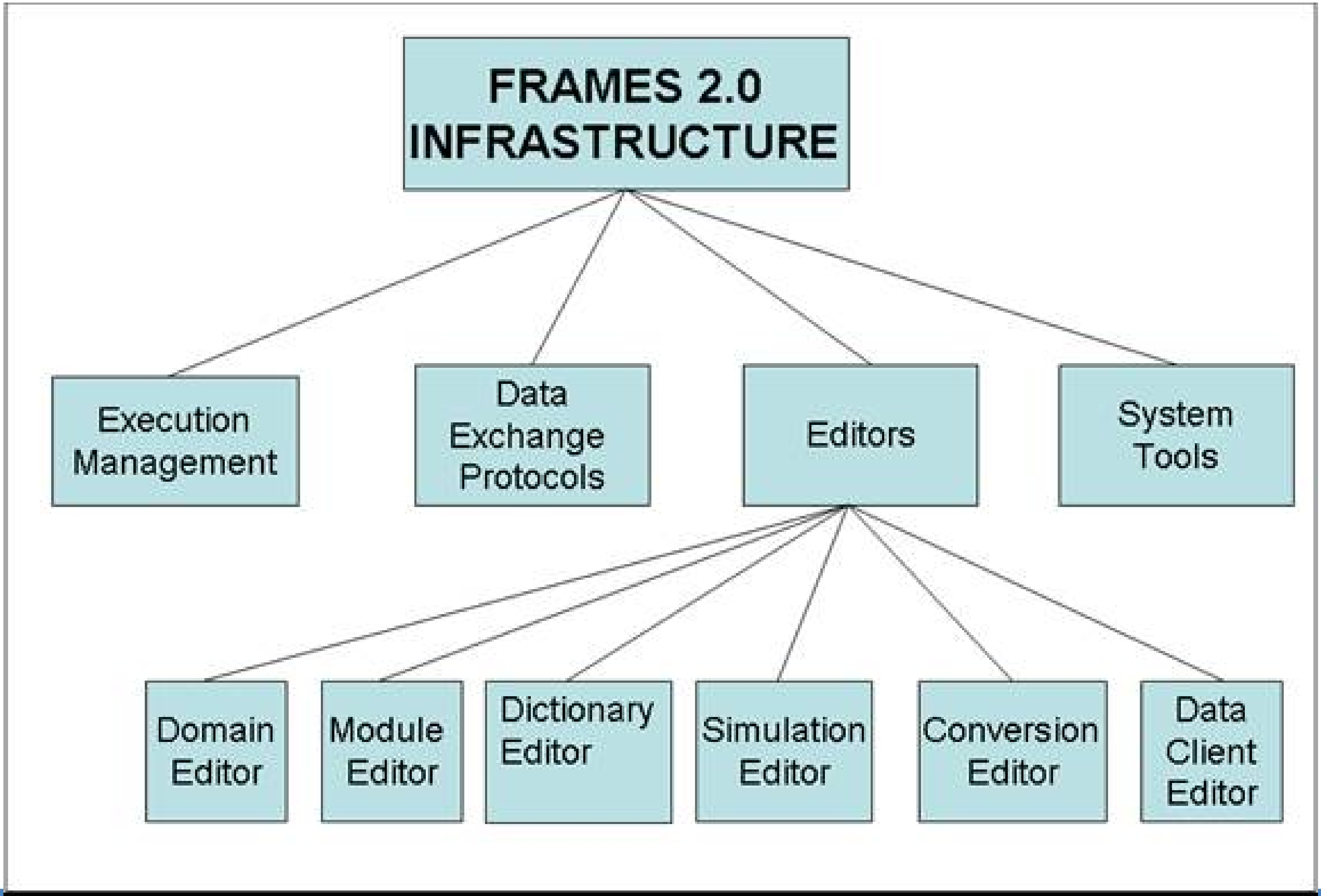
D4EM : **(Data for Environmental Modeling)**

a software system for accessing, retrieving, and processing (including Geo-processing) of data for integrated modeling systems

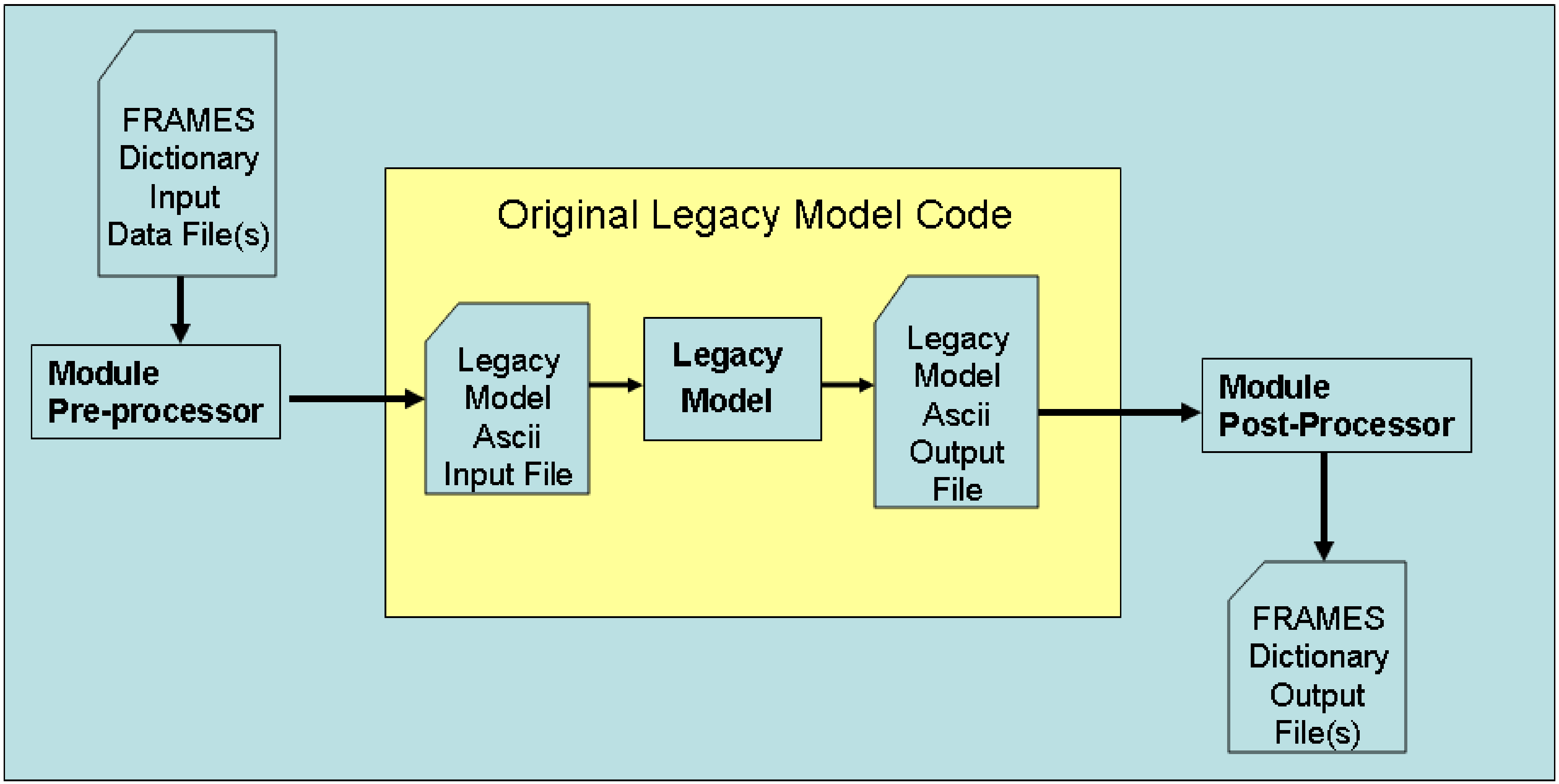
SuperMUSE : **(Supercomputer for Model Uncertainty and Sensitivity Evaluation)**

a software system that facilitates the execution of FRAMES based modeling systems across a clustered network of PCs

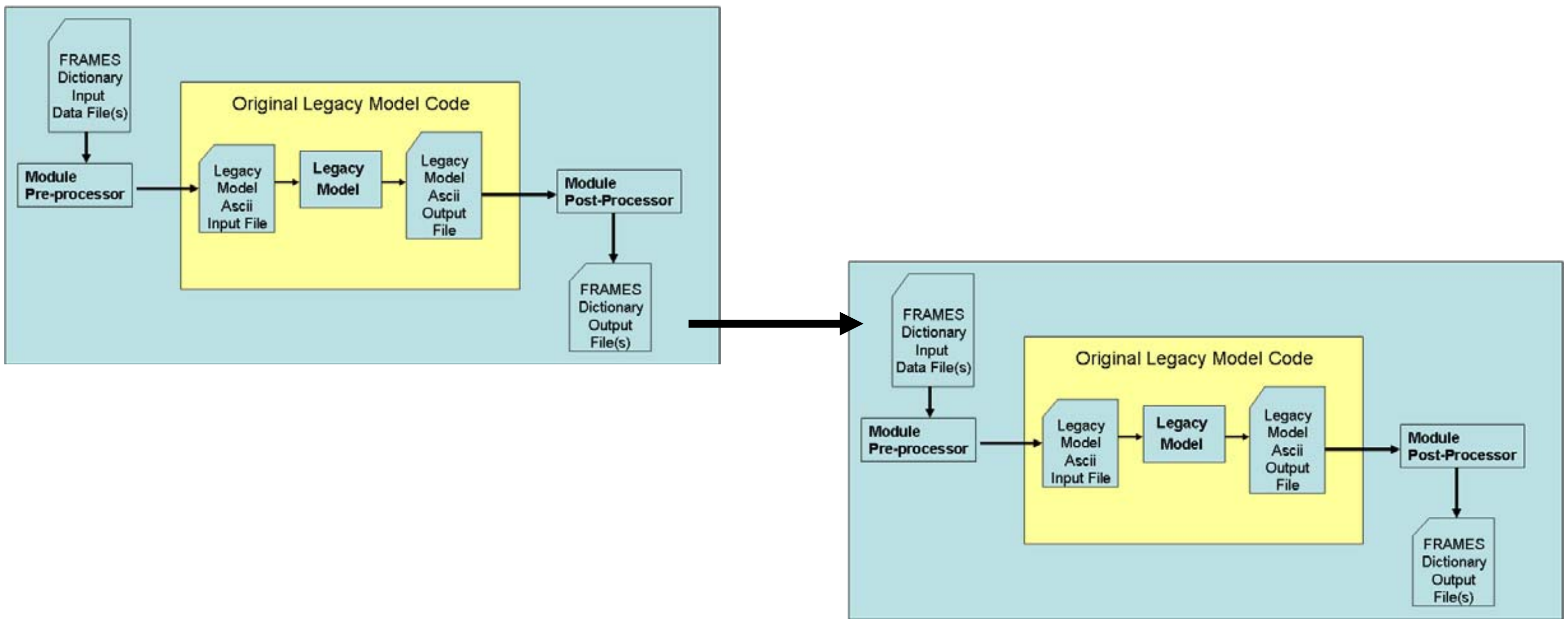

Frames Development Environment - [Simulation Editor - MEERT Domain]
File Editors Tools Customize GO... Help Refresh...
FRAMES
Database
Model
Fate and Transport
Aquatic Ecosystem
Landscape
Chemicals
Climate
EcoServices
System
Unit Definition
D4EM
Model Evaluation Compor
Iterators
Samplers
Summarizers
Constructed Triplet
Parameter Estimation
Single Solution
Multiple Solutions
Viewers
Viewer
Model Output
UDP (Mod1)
SWAT (Mod2)
WatershedHg (Mod3)
WASP (Mod4)
HSI (Mod5)
BASS (Mod6)
esp (Mod9)
Model Out
File: c:\program files\framesv2\meert\simulations\meert_apes.sim
Icon: Mod1
10/9/2009
3:01 PM


FRAMES 2.0
INFRASTRUCTURE
Execution
Management
Data
Exchange
Protocols
Editors
System
Tools
Domain
Editor
Module
Editor
Dictionary
Editor
Simulation
Editor
Conversion
Editor
Data
Client
Editor

FRAMES Assimilation of Legacy Models

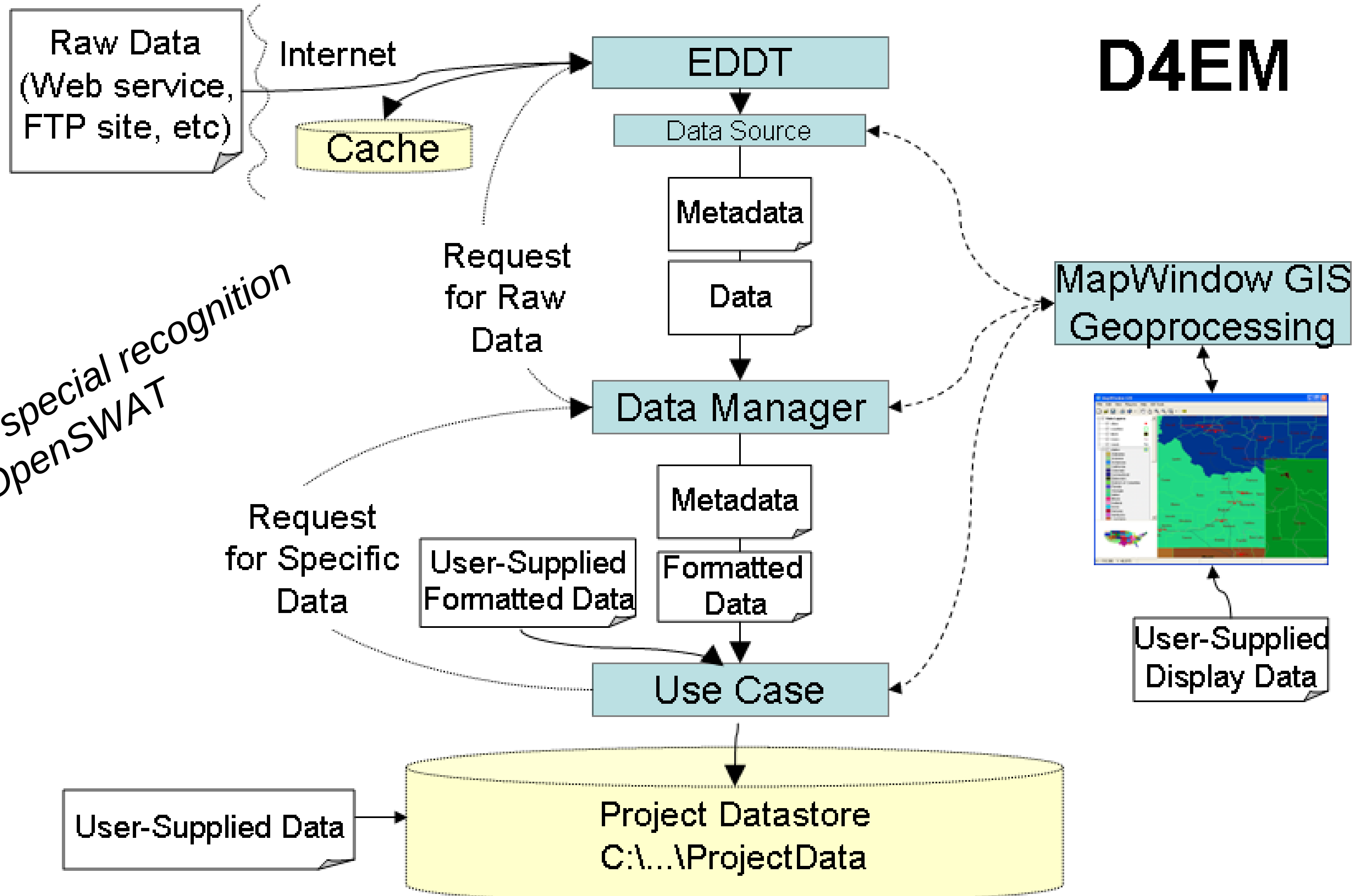


Simplified view of Model Linking in FRAMES

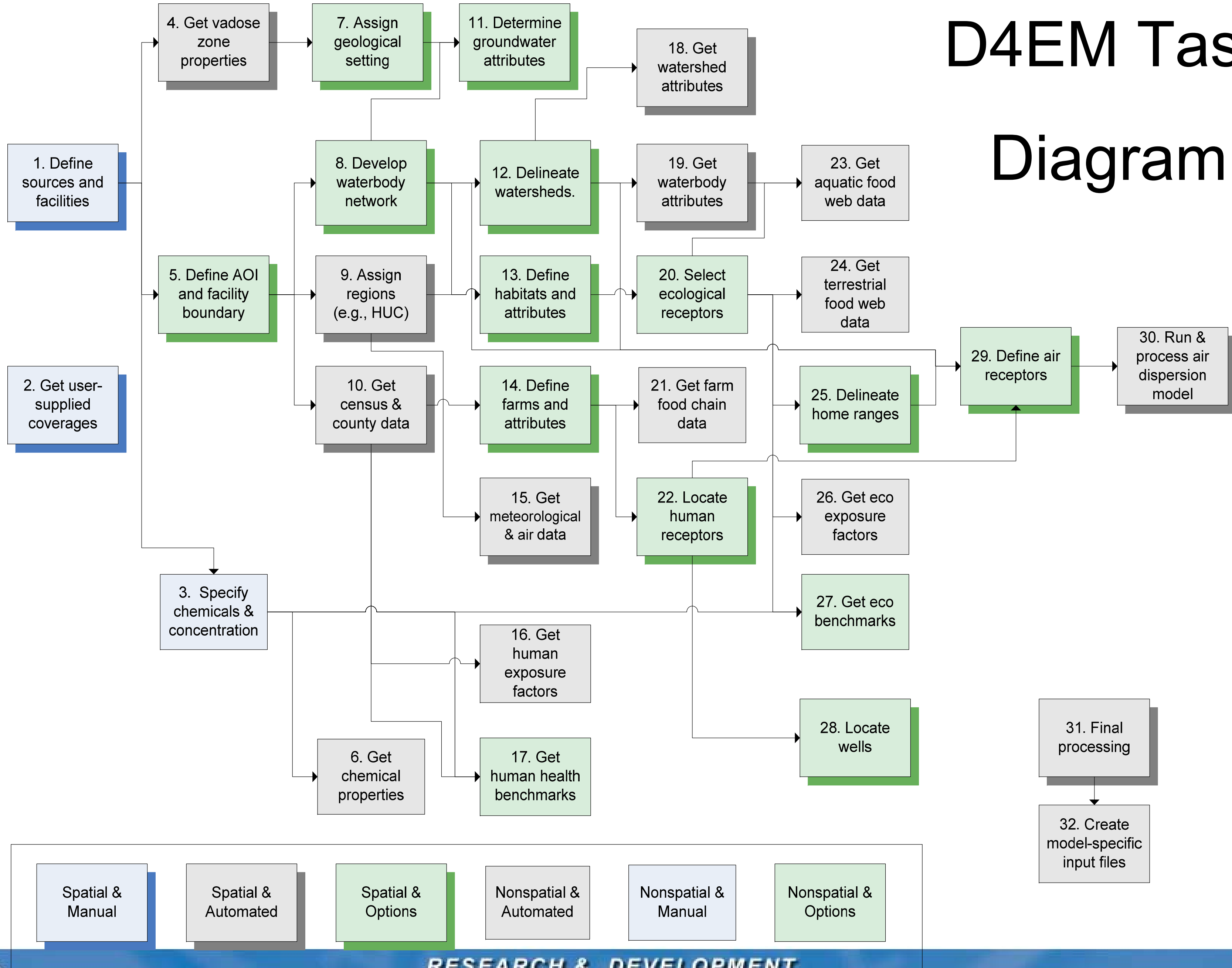


Data for Environmental Modeling

With special recognition of OpenSWAT



D4EM Task Diagram

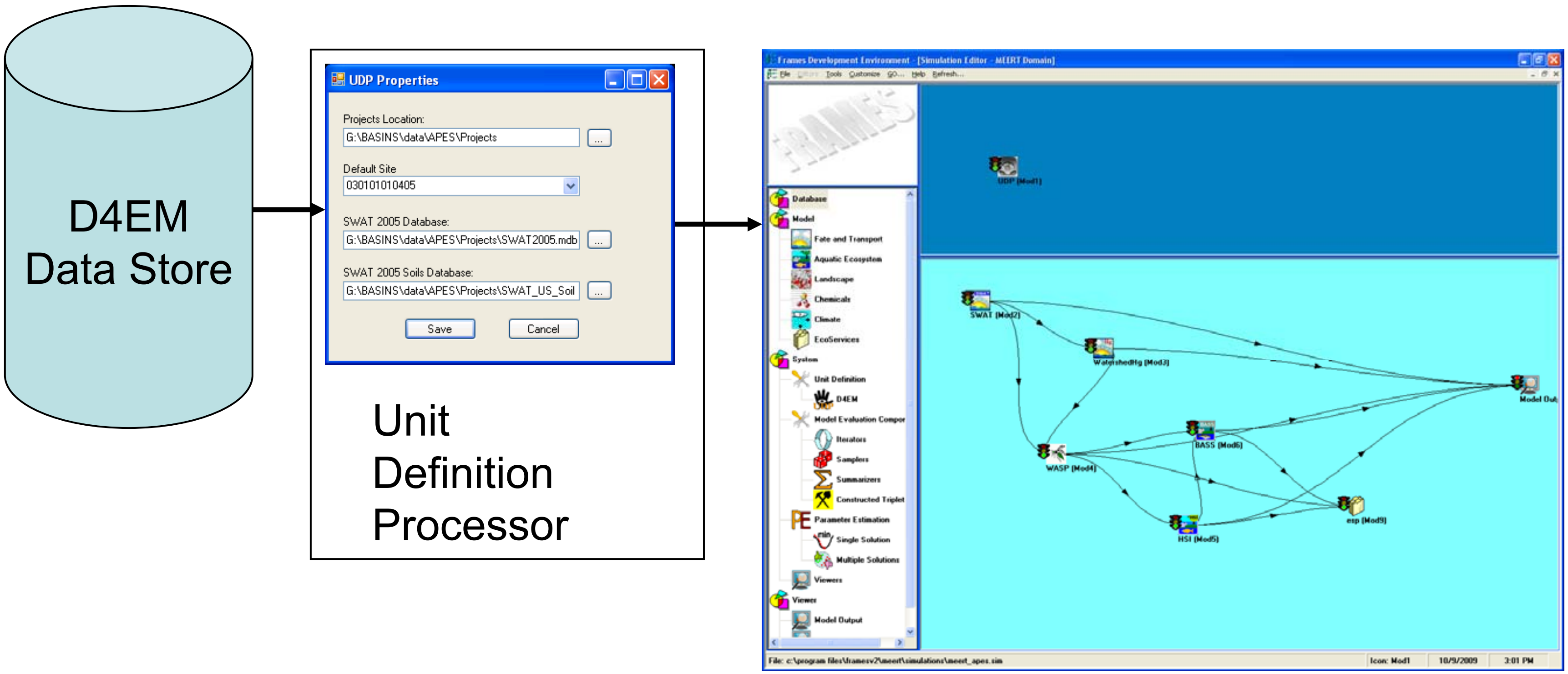


D4EM Data Sources

BASINS
- Land use/land cover
- Urbanized areas
- Populated place locations
- Reach File version 1 (RF1)
- Elevation (DEM)
- National Elevation Dataset (NED)
- Major roads
- USGS HUC boundaries
  - Accounting unit
  - Cataloging unit
- Dam sites
- EPA regional boundaries
- State boundaries
- County boundaries
- Federal and Indian lands
- Ecoregions
- Legacy STORET
- STATSGO
- MET Data

- NLCD
- NWIS
- NHDPlus

Transfer Data from D4EM Datastore to Modeling System



SuperMUSE Parallel Computing Cluster

3MRA Version 1.x

SuperMUSE – Supercomputer for Model Uncertainty and Sensitivity Evaluation

The Integrated Team

- **Ecology**
  - Craig Barber
  - Brenda Rashleigh
  - Tom Purucker
  - Mike Cyterski
  - John Johnston
  - Luis Suarez
  - Dan McGarvey
- **Watershed Hydrology & Erosion**
  - Steve Kraemer
- **Watershed Chemical F&T**
  - Chris Knightes
  - Bob Ambrose
  - Steve Kraemer
  - Heather Golden
- **Surface Water Quality**
  - Bob Ambrose
  - Chris Knightes
- **Atmospheric Deposition**
  - Ellen Cooter
  - Robin Dennis
- **Multi-media Model Integration**
  - Gene Whelan
  - Gerry Laniak
- **Farm Practices**
- **Uncertainty Analysis/Regional Roll-up**
  - Justin Babendreier
  - Tim Shaw
  - Siddharth Sharma
  - Karl Castleton
- **Data Processing and Modeling Infrastructure**
  - Kurt Wolfe
  - Rajbir Parmar
  - Jack Kittle
  - Mark Gray
  - Mitch Pelton
- **GIS Processing Support**
  - Lourdes Prieto

Concluding Remarks

- **Modern Integrated Environmental Modeling is critical to Decision Making**
  - It involves a high degree of cross-disciplinary science and communication (data, process knowledge, models)
  - It is conducted across spatial scales ranging from local to regional to national to global and across temporal scales ranging from seconds to years to decades
  - It is “systems” oriented
- **Modern Software based technologies are critical to Modern Integrated Modeling**
  - They involve the large scale integration of and communication among data, models, methods, and humans)
- **Integration of interdisciplinary science, people, and technology is not rocket science, it’s way more difficult**
  - It takes time, patience, and a high degree of communication