DHARMA:
Domain-specific Metaware for Hydrologic Applications

1 Objectives and Significance of Proposed Activities

We intend to build a middleware layer to provide the resources needed to revolutionize hydrologic modeling. The required resources range from local data to the supercomputing power on the national computational Grid.

Our objective is to advance the fields of hydrology and computer science. In hydrology, we are proposing to expand the applicability of the WEPP (Water Erosion Prediction Project) model and the SITES (Water Resource Site Analysis) model to large watersheds, specifically applying the extended model to the Lake Decatur watershed in Illinois. We intend to enable the development of new models for predicting erosion within watersheds by allowing significantly easier access to the computational power and data acquisition capabilities of the Internet. Within computer science, we intend to advance the state of the art in automatic distributed data acquisition and distributed scheduling, utilizing the power of recent developments in metacomputing (e.g., Globus and Legion) and digital libraries (e.g., the Alexandria Digital Library at UCSB).

1.1 Our goals

We plan to develop a layer of middleware to interface several hydrologic simulations with the computing power of the National Computational Grid, and the domain-specific informational resources available on local workstations and through the Internet. The middleware layer, DHARMA, has the following major points of functionality:

- automatic data acquisition via the Internet for many types of data, in particular, geotemporal climatic data from online databases and digital libraries; merge the data necessary for the computation to occur, from online, local, and cached resources,
- smart caching of intermediate results to allow for reuse in future simulation cycles,
- acquire a task graph from the UI layer, and optimize it through application-specific knowledge and dynamic results caching strategies; optimize execution times through awareness of remote and local resources, including memory, processor, storage, and data, and
- interface with computing resource managers, such as Globus and Condor.

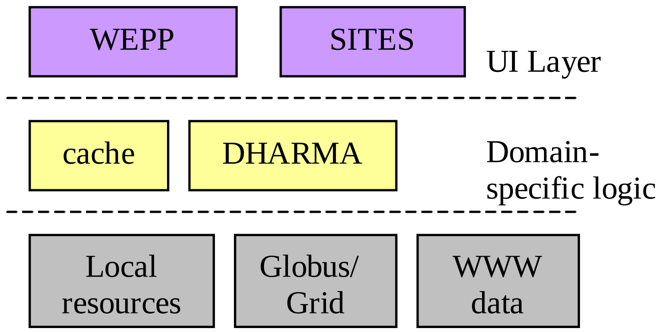

Figure 1: DHARMA block diagram.

Major technical problems to be solved include: How do we automatically acquire the necessary data? How do we know what items can and cannot be cached? How can results be passed from simulation to simulation with dissimilar data formats? What DTD's must be developed for information within this domain? What domain-specific task graph heuristics are needed for efficient computation?

To address these issues, we plan to:

- use XML[1] and DTD's[2] as a lingua franca between simulations, remote data sources, and DHARMA components.
- realize our primary goal is flexibility and ease in data acquisition and data transfer between simulations, which cannot be sacrificed for utility-limiting performance enhancements.
- allow DHARMA to have multiple interfaces to existing metacomputing system. Use DHARMA to package up the computation into a tidy set for computation by these systems.
- develop heuristics for scheduling task graphs, interfaces for the various resources, and application-specific caching techniques.

The system is intended for use not only by hydrologic researchers, but also by engineers in the field working on a day by day basis. These engineers typically have older, inadequate local computing power for the increasingly complex models required. Use of the DHARMA interface will allow the engineers to tackle problems on a scale impossible without sophisticated domain-specific computational management systems.

The WEPP watershed model is a continuous simulation processes-based model which represents new soil erosion prediction technology based on fundamentals of stochastic weather generation, infiltration theory, hydrology, soil physics, plant science, hydraulics, and erosion mechanics. The applicability of the model, however, is limited to very small watersheds (up to few hundred acres) due to currently available computer technology in handling input data requirements. The model is capable of estimating spatial and temporal distribution of soil loss, runoff, and sediment yield, and can be extrapolated to a broad range of conditions that may not be practical or economical to field test. WEPP is the only model that is currently

[1] eXensible Markup Language
[2] Document Type Definition

available for accurate prediction of soil erosion and sedimentation based on fundamental scientific theory and principles.

If the WEPP model can be advanced so that it can be applied to actual watersheds that are often several thousand acres, *it will bring a revolutionary change* in hydrologic modeling on the watershed scale. Currently, all watershed scale hydrologic models are based on empirical relationship and seldom include process interactions among soils, hydrologic, plant, and atmospheric processes. WEPP is the only model which accounts for soils, sediment transport, runoff, channel flow, plant growth, decomposition, snow melt, freeze-thaw effects, and climatic conditions. If this model can be enhanced for its applicability over larger watersheds, it will be the only such model which will provide accurate predictions and evaluate effects of alternative watershed management practices on watershed water quality. Developing the sophisticated automated data acquisition tools to support this endeavor will require substantial advances in distributed algorithms, which will in turn spur advances in other areas.

1.2 Justification and strategy for achieving goals

Building on previous work In our recent research, we have focused on high-performance distributed computation over the Internet. This has included developing scalable WWW servers, a hierarchical scheduling system for HTTP[3] servers, and intelligent scheduling strategies for software agents [1, 2, 10, 3].

Currently, we are developing partnerships to explore the automated acquisition of scientific data from the Internet, building on the numerous “wired” databases and digital libraries across the world. We have partnered with the Wind Erosion Research Unit (WERU) at Kansas State University to use various Internet standards and create a particle pollution prediction tool (discussed in Section 2). We are working with Kansas Water Resources Research Initiative laboratory at KSU to combine soil moisture simulations with online weather information to provide a web-based irrigation advisory tool for Kansas farmers.

We are also investigators on the Alexandria Digital Library (ADL)/ Alexandria Digital Earth Project (ADEPT) High Performance and Parallel Processing Team at the University of California, Santa Barbara. ADL and ADEPT, as NSF DLI and DLI2 projects, involve research on issues critical for the construction of distributed digital libraries of geospatially-referenced, multimedia materials. We have developed several mechanisms for enhancing the scalability of Web servers in the digital library environment. Our initial efforts offered the first content-based request distribution across a network of workstations (NOW) acting as a single virtual server [1]. We then introduced the concept of the *task chain*, which allowed the server cluster to dynamically schedule request processing across the servers *and* make use of client resources [2, 4]. Making use of client resources and multiple servers gave speedups of more than 3000% over round-robin scheduling, and improved load balancing. We are now exploring ways to go beyond HTML and utilize the immense data stores in the digital libraries while retaining their structure and metadata.

We have also recently developed general scheduling mechanisms utilizing our sophisticated algorithms matched with custom and general metacomputing system resource discovery mechanisms. Our first application was a hierarchical Web scheduling system, called HSWEB, which organized clusters of servers into a unified hierarchy [3]. The system considered multiple aspects for scheduling purposes, including cached dynamically created data, the costs of moving data vs. moving computation, the amount of client

[3] HyperText Transfer Protocol

CPU available, and the bandwidths throughout the system [3]. Utilizing the additional resources of the hierarchy wisely gave improvements of over 250% in several applications. We have also applied the same technology to software agents. We combined the HSWEB scheduling infrastructure with IBM's "Aglets" Java-based mobile agents, which yielded an improvement of over 250% in our sample application [10, 13].

Strong regional and national collaborations Each of the above advances was made possible through strong regional and national collaborations. Our scheduling infrastructure relies heavily on the Network Weather Service (NWS) for cluster-to-cluster bandwidth and latency information [24]. The NWS was developed as part of the AppLeS project at University of California, San Diego (UCSD), and is part of the Globus metacomputing project [8]. We are in a continuing collaboration with its developers at UCSD, as well as Dr. Wolski, who is on the faculty at the University of Tennessee, Knoxville, to improve and utilize its capabilities.

Access to the facilities and people of the Alexandria Digital Library Project made possible our work on scalability in their environment, and they have committed continued access and cooperation to our joint exploration of ways to make their work more accessible and useful. Our driving applications for the proposed project are from regional laboratories, including Water Resource Sites Analysis Software (SITES) by Mitchell L. Neilsen (KSU), Darrell M. Temple (USDA), and others; WEPS from the Wind Erosion Research Unit (WERU) at Kansas State; and data collaboration by Prasanta Kalita, of the University of Illinois at Urbana-Champaigne (UIUC) [18, 25, 12].

2 Background

What we are trying to do, technologically, is to present the researcher with an intuitive way to link existing simulations. By drawing the desired graph of operations to be performed (see Figure 2), the system would have the necessary information to automatically extract the required data from the distributed data servers, run the simulation(s), and present the results.

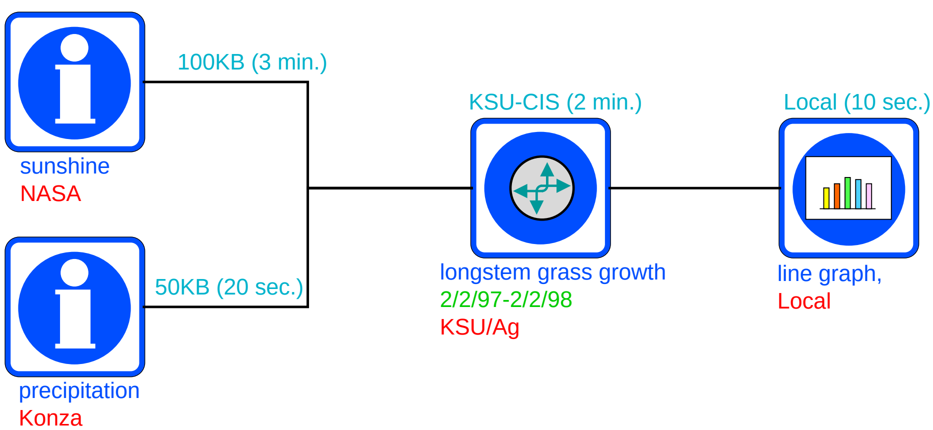

Figure 2: A simple calculation specification.

Presenting such an interface to the researcher and educator, with its seamless and transparent access to distributed databases, is a daunting technological challenge. Various other distributed computation systems have previously been developed but almost invariably are aimed at automating the distribution of existing scientific programming and data. Systems such as AppLeS, Globus, and Legion all will schedule computation quite competently [5, 8, 11]. However, they fail to address the most fundamental difficulty in conducting any sort of operation on distributed data sets – the acquisition and matching of data to the models being used. Furthermore, the interface presented to the user is often more suited to the professional programmer rather than a researcher in the physical sciences. Systems devoted to universal data access, such as the WebHLA and DATORR projects, lack the domain-specific knowledge to acquire and optimally use the necessary hydrological data [9].

2.1 A prototype

We have developed a simple prototype in the area of particle pollution prediction. Most particular pollution is caused by wind-caused soil erosion, for which sophisticated simulations are available. One such, the Wind Erosion Prediction System (WEPS), is under development at WERU. We plan to utilize weather forecasts, a historical weather database, WEPS, and various heuristics to give a near-term prediction for particle pollution levels.

The overall goal is to develop a prototype PM10[4] warning system with one to two day areal forecasts for high PM10 levels from wind erosion in the state of Kansas. Our prototype initially automates the collection of the weather data using the Internet. The prototype, depending on the date ranges desired, collects historical weather information from the pertinent weather stations and obtains forecast weather information from the NOAA WWW site. It transfers this data via XML to a “wrapper” application, which formats the generic weather information into a seamlessly merged set of input files for the WEPS simulation. WEPS then produces the pollution prediction.

Historical weather information is available from 1986 to the present. This data is from the KSU Automated Weather Stations network, consisting of 15 stations scattered across Kansas providing hourly information (uploaded to the central store daily). The automated sites provide temperature, wind, solar radiation, and ground temperature information.

We have chosen an XML-based data format for several reasons. Using XML leverages a huge wave of support from throughout the industry in terms of tools and source code. However, the primary advantage is the combination of device- and OS-independent data transfer, with the ability to retain its semantic structure [26, 6]. This is a critical advantage over text files or HTML. Unlike native TCP/IP sockets, the XML utilities allow the easy transfer of structured data. Furthermore, since the semantic information is retained, a single information source can serve multiple applications. For example, our weather data server can easily be used by other weather-based simulations.

We have developed a DTD for weather information in conjunction with WERU and other environmental scientists at KSU, a fragment of which is shown in Figure 3. The structure of the data is easily human-readable, and tools automatically create structured data from documents conforming to this standard. We have extended the basic elements with a comment field, in which we plan to list the transformations the data in a particular field has undergone. For example, noting a conversion from Fahrenheit to Celsius with 8-bit precision, or that the data was generated by a particular simulation, can give valuable clues about the accuracy of subsequent analysis.

Alternative technologies to XML include CORBA, sockets, NetCDF, Java RMI, or other custom networking protocols. CORBA typically is more complex for developers, but provides an object-oriented

[4] particles of pollution 10 microns or less in size

```
<?xml version="1.0" encoding="ISO-8859-1"?>

<!ELEMENT weather       (record)*>
<!ELEMENT record        (days, data)>
<!ELEMENT source_id     (#PCDATA)>
<!ELEMENT comment       (#PCDATA)>
<!ELEMENT days          (day_id, when, comment*)>
<!ELEMENT day_id        (#PCDATA)>

<!ELEMENT when          (day, month, year, comment*)>

<!ELEMENT day           (#PCDATA)>
<!ELEMENT month         (#PCDATA)>
<!ELEMENT year          (#PCDATA)>

<!ELEMENT data          (hourly_info | daily_info | abnormal_data)>

<!ELEMENT hourly_info   (data_id,    comment*,
                                source_id?,  comment*,
                                day_id,      comment*,
                                hour_min,    comment*,
                                temperature, comment*,
                                rel_humid,   comment*,
                                wind_dir,    comment*,

  ...
```

Figure 3: A subset of the weather DTD.

interface [21, 22]. NetCDF is designed for the exchange of scientific information, and offers an intrinsic Web interface, but has failed to achieve the same level of support across the entire Web industry as XML [15]. We hope to provide a gateway between our system and NetCDF data sources, due to the significant amount of scientific data available in that format. Java RMI provides an excellent mechanism for communicating between Java processes, but lacks the flexibility and genericity of XML [23]. Through the use of industry-standard information exchange mechanisms, we can provide for future expansion and compatibility with online information sources.

3 Hydological watershed models

We plan to adapt WEPP and SITES, from the hydrologic arena, to our distributed computation and data acquisition system. Many simulations at the watershed scale are limited in resolution and scope by their computational demands. Distinguishing features of these simulations include large-scale model and data integration requirements (many small models with differing data requirement), the need for specific resources (including semi-real time data acquisition) which are not within the management scope of the national Computing Grid, and significant opportunities for domain-specific optimizations such as caching and data reuse.

3.1 WEPP

The WEPP watershed model is a continuous simulation processes-based model which represents new soil erosion prediction technology based on fundamentals of stochastic weather generation, infiltration theory, hydrology, soil physics, plant science, hydraulics, and erosion mechanics. The model has been developed under the leadership of the USDA-Agricultural Research Service with cooperation from USDA-Natural Resources Conservation Service, USDA-Forest Service, USDI-Bureau of Land Management, and other organizations involved in soil and water conservation and environmental planning and assessment. The model is capable of estimating spatial and temporal distribution of soil loss, runoff, and sediment yield, and can be extrapolated to a broad range of conditions that may not be practical or economical to field test. WEPP is the only model that is currently available for accurate prediction of soil erosion and sedimentation based on fundamental scientific theory and principles. The applicability of the model, however, is limited to very small watersheds (up to a few hundred acres) due to currently available computer technology in handling input data requirement. The fundamental processes considered in WEPP are soil erosion, sediment transport, and deposition, infiltration, residue and canopy effects on soil detachment and infiltration, surface sealing, erosion hydraulics, surface runoff, plant growth, decomposition of residue, percolation, evaporation, transpiration, snow melt, frozen soil effects on infiltration and erodibility, climate, tillage effects on soil properties, and the effects of conservation practices on soil; all of these processes are considered with spatial and temporal variability in topography, surface roughness, soil properties, crops, and land-use conditions.

Since the soil, biological, hydrologic, and atmospheric processes interact with each other on spatial and temporal basis, the model requires a significant amount of input data and computations, which limit the applicability of the model to watersheds that are very small. Nevertheless, WEPP is the most dependable model for accurate prediction of soil loss and sedimentation. If a new input data handling structure can be built and a new computational method developed for the spatially dependent interactive processes, the model may be used for larger watershed and will bring a revolutionary change in hydrologic computational processes.

3.2 SITES

Vegetated earth (soil and rock) auxiliary or emergency spillways have been used extensively on reservoirs and ponds for flood control within the United States. These spillways generally consist of a trapezoidal channel cut through natural materials and vegetated as appropriate for the local area [16]. The USDA Soil Conservation Service (SCS) has constructed approximately 23,000 structures using this type of spillway [7]. Despite their widespread use, the processes by which these spillways erode during extreme events, and their effect on the environment, are only imperfectly understood. For many years, research utilizing both laboratory and field data has been conducted to improve criteria for design and analysis of individual spillways.

SITES software evolved from the DAMS2 software, which has been used extensively by the Soil Conservation Service for the design and analysis of flood control structures since 1971. While SITES retains all of the capabilities of the DAMS2 software, it also adds a number of new features to better predict erosion. For example, it has been found that erosion of vegetated earth spillways can be divided into three phases: vegetal cover failure, concentrated flow erosion, and headcut advance [16]. Hydraulic and hydrologic analysis of structures (dams) designed using this new NRCS criterion for spillways and flood routing can

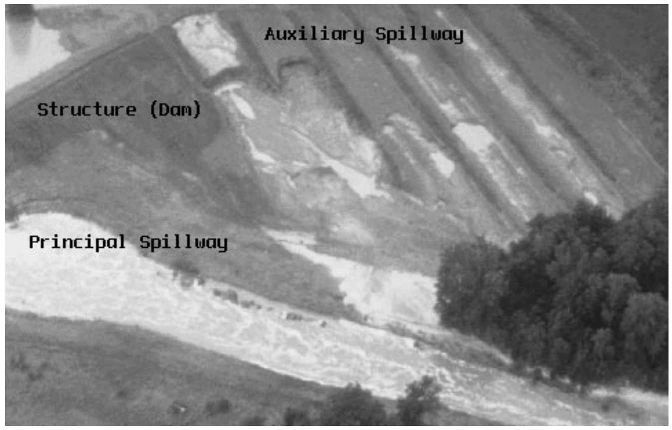

Figure 4: Spillway erosion.

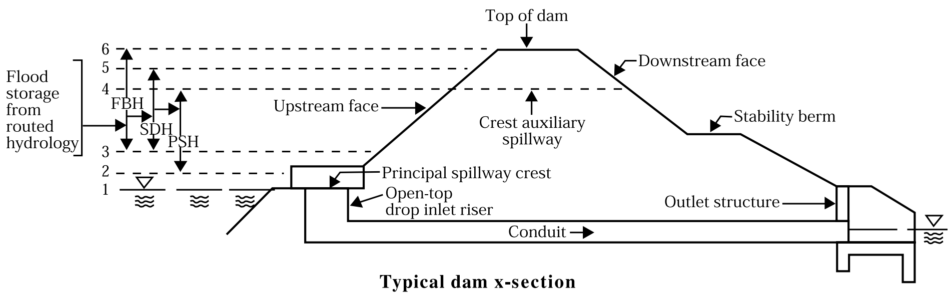

Figure 5: SITES Dam components.

be performed using SITES. The SITES software was developed using Fortran, and consists of several thousand lines of source code [18].

Input to SITES includes information about the structure being analyzed and information required to generate the inflow hydrograph(s) – functions that represent the amount of water flowing into the reservoir created by the structure (see Figure 5). Hydrographs, or the rainfall/watershed information required to generate them, may be derived from historical data for a given flow event, or generated automatically from standard design tables. Execution of SITES generates outflow hydrographs and other structural information, including predicted earth spillway erosion. This information is stored in several ASCII text files.

SITES has recently been upgraded to include a graphical user interface to facilitate data entry for analysis of a single site on a wider array of computer systems, including personal computers [17]. However, there is still a need for procedures to better predict spillway performance and safety at the watershed scale, and to provide end users with efficient design tools. In this project, we have designed an integrated development environment (SITES-IDE) that can be used to analyze the technical implications of various hydrologic or hydraulic conditions on a given spillway design, and to efficiently compare alternative spillway designs. The integrated development environment also allows a designer to exploit the parallelism inherent within the application.

Figure 6: Spillway breach.

3.2.1 Earth Spillway Design

Spillway designs are compared by determining both spillway stability and integrity when the spillway is subjected to a given design storm.

In a typical design, three types of inflow hydrographs are used: principal spillway hydrographs, stability design hydrographs, and freeboard hydrographs.

The *principal spillway hydrograph* is used to size the principal spillway and set the elevation of the crest of the emergency or auxiliary spillway. The principal spillway is typically a conduit through the dam used to pass low flows, whereas the auxiliary spillway is often an open channel capable of passing infrequent large flows. Earth auxiliary spillways are typically wide trapezoidal channels vegetated as appropriate for the local area.

The *stability design hydrograph*, when routed through a reservoir, generates the maximum auxiliary spillway outflow that the reservoir will be expected to pass without erosion damage. For the design to be stable, erosion thresholds must bound hydraulic stresses. The erosion threshold is the stress level associated with the initiation of erosion on the spillway. For flows larger than the stability design hydrograph, spillway erosion may occur. As shown in Figure 4, the eroded spillway may require maintenance [19].

The *freeboard hydrograph* represents the maximum flow for which the structure is designed. The integrity of the auxiliary spillway, as represented by its resistance to breach, is evaluated for the spillway outflow associated with this hydrograph. This is the most important consideration in designing an earth (soil, rock, or both) spillway. Even though extremely large discharges may cause significant erosion, the spillway must not breach during passage of the *freeboard hydrograph*. A spillway is considered breached if the spillway crest is degraded by erosion and floodwater is released through the spillway below the crest elevation.

Figure 6 shows an example of a breached spillway. Breach potential is a function of the spillway system, the characteristics of the spillway outflow hydrograph, the erodibility of the earth materials, the spillway layout, bottom width, and maintenance. Open channel earth spillways must perform satisfactorily both hydraulically and structurally. From a hydraulic standpoint, the spillway must handle the required range of flows with a predictable stage versus discharge relationship. From a structural standpoint, the spillway design must consider both stability and integrity. The analysis is based on the idea that some erosion is allowable if its occurrence is infrequent, maintenance is provided, and the spillway will not breach during passage of the freeboard hydrograph [18].

The integrated system for water resource site analysis (SITES-IDE) is designed to fully integrate the simulation model (SITES) within a graphical development environment. It interactively guides user input, invokes the simulation model in the background, and parses the results to automatically generate output hydrographs, summary tables, and graphs. The system provides tools that can be used to better understand the structure, function, and dynamics of water resource control structures.

It also provides a real-world application that can be used to experimentally evaluate new parallel and distributed systems. The code for both the simulator and integrated development environment are readily available.

In this project, we plan to fully incorporate topological models and allow those models to be generated automatically from a geographic information system (GIS). We also plan to provide canonical hydrograph representations to allow efficient integration with other watershed modeling software. For example, a hydraulic engineer may want to use another model, such as TR-20 [20], to route hydrographs down a stream, and SITES to evaluate the vegetated earth spillway at a site.

Another interesting issue is exploitation of parallelism and element interdependencies between runs. After some elements have been modified, it is easy to automatically determine which hydrologic elements need to be recomputed. In addition, related runs will typically contain numerous elements in common. It is possible to model all elements as objects and maintain dependency information between elements and runs. Then, changes made to any hydrologic element will be automatically reflected in all runs.

4 Research and Education Plan

4.1 Overarching objectives and strategy

We briefly review the foundational and supporting objectives of the project, and suggest concrete results by which the success of the project should be judged. This is followed by a more detailed discussion of our strategy for realizing the objectives.

Foundational objective: The cornerstone of our project is the construction of a reliable system for linking data sources simply and correctly. Data sources may be local or remote, static or dynamic. This system *must* be present to serve as a foundation for experiments and refinement of methodology. Realizing this objective is primarily a matter of *technology,* can existing techniques be scaled to handle large and diverse software systems?

Supporting objectives: The two other primary measures of success for this project are the *validation of system philosophy* and development of the *educational infrastructure*. Both of these are discussed below.

4.2 A full-scale operating environment

> *Our basic strategy is to take our prototype's XML-based information transfer abilities, and add our domain-specific application interface to existing distributed task scheduling/load balancing abilities.*

A typical watershed being simulated by our system might consist of several subwatershed objects, each simulated by a different model. Running these individually, and manually transferring the results as needed to other simulations, is a painstaking and time-consuming task. This leads to the objectives within this phase.

Distributed object model The first objective is to develop an *extensible, object-oriented framework* for representing elements such as simulation models and data. The current prototype does not provide a model for describing a dependency graph of models and data sources. At the top level, simulation models are encapsulated as objects requiring certain resources in terms of input data and computational resources. A complex simulation may involve interaction between several simulation models. At a lower level, hydrologic components in the watershed (cells, streams, reservoirs, etc.) are encapsulated as object with properties defining their attributes (vegetal cover, slope, elevation, etc.) and methods defining operations on the objects (surface flow, infiltration, etc.). Some attributes may have spatial or temporal distributions, while other attributes may not. By defining a canonical representation for various types of objects via a class-specific DTD, we will be able to more easily combine complex simulations in a holistic system.

Integration with GIS The second objective is to develop a *fully integrated spatial tool* that provides the linkage between a geographic information system (GIS) and the simulation models. By using this tool, the user will be able to specify the spatial extent for each simulation model, and the relationships between hydrologic elements within the model. Typically, spatial attributes for those objects can be derived from existing spatial layers stored in the GIS. A user can associate other information with a simulation object by double-clicking on the object or by selecting and clicking on a tool bar icon. In either case, a separate user interface may be invoked to specify non-spatial data for certain items such as simulation-specific parameters. After all data required for a given simulation model has been entered, the user will be able to invoke the simulation model and visualize the results using the same spatial tool, along with the rich set of visualization tools provided by the GIS. As necessary, we will extend the capabilities of current simulation models so that they may be fully integrated. For example, many simulation models have been developed using Fortran 77, and only run under MS-DOS. These models will be recompiled using a 32-bit compiler to run under NT or Unix. We will define a standardized method for encapsulating simulation models so that they all provide a canonical interface using Java wrappers and XML technology. We have extensive experience in developing similar GIS tools for sustainable farm management practices [14].

For example, a typical simulation for many watershed models involves routing a set of hydrographs through the hydrologic elements within the watershed. However, the data formats for input and output are not standardized. A hydrograph is a function that represents the amount of water (discharge) flowing from a hydrologic element over time. This function can be represented using several different formats:

1. a set of (time, discharge) coordinate pairs,
2. a single starting time and delta time, along with discharge values, or
3. a set of (time, cumulative discharge) coordinate pairs.

Furthermore, the position and precision of the time and discharge values also vary. We hope to infer the types and appropriate sources for data from the simulation specifications, which will simultaneously limit the amount of user interaction required, and allow several key optimizations. First, we can apply data filtering at the source site, potentially substantially reducing the amount of bandwidth (and hence time) needed. Second, we can perform limited transformations automatically, such as converting a table of observations from integer to floating point numbers.

Human resources The implementation will be headed up by the principal investigators. Several M.S. students will also assist in the development throughout the duration of the project. We will be aided by interaction with our collaborators where necessary for application-specific requirements.

Time table and concrete results

Years 1-2: build basic system functionality
Year 2: build GUI and integrated system with GIS

4.3 Validation of system philosophy

Our firm belief is that this system has the potential to be a great asset to scientific researchers. We must evaluate the system's effectiveness. In this evaluation, we will consider its effectiveness as a research tool; that is, its effectiveness in providing a faster, more convenient means to access and utilized distributed simulations and data. The development of these large-scale models will be a joint effort between our collaborating scientists at WERU and the USDA and ourselves.

Evaluation via large systems Biological and environmental scientists and engineers will further develop and utilize the system for *modeling complex hydrological systems over larger geographic regions*. Current techniques do not allow accurate simulation of the physical and hydrological processes primarily due to limitations on data availability, inadequate representation of process equations by the models, and the limited ability of model prediction for larger areas. The proposed system will allow accurate temporal and site-specific predictions of environmental and hydrologic parameters. This will provide scientists and engineers with a tool to design and develop best management practices to protect natural resources and the environment. In addition, scientists and engineers can use the system to develop and evaluate new environmental models which are composed from existing models.

A major breakthrough will be the advancement of the WEPP capability itself, which will be restructured for parallel execution, and extended with new behavioral equations. The model, which has been developed by a team of scientists from multiple federal organization, will be applied for accurate prediction and environmental assessment. Once we improve or advance the model to that level, we will test its performance on Lake Decatur watershed. Since we have about 30 years of data collected over this watershed, we will be able to calibrate and validate the model. Once validated, this model can be used for environmental assessment for any watershed in the country.

The Lake Decatur watershed comprises of 925 square miles and provides the public water supply for the people and industries of the City of Decatur and the Village of Mt. Zion, Illinois. Most of the watershed land is privately owned farmlands, used primarily for corn and soybean production. The circumstances contribute to a large number of water quality problems including high nitrate-N, sedimentation resulting in a gradual loss of reservoir storage capacity, and turbidity (high sediment concentration in water) at levels high enough to cause water treatment problems. The city is working with the Illinois State Water Survey (ISWR) to perform hydrologic and water quality monitoring. The ISWR has used modeling to evaluate the effects of current land-use practices on water quality and sedimentation problems; however, the empirically based models have not provided much useful information. If WEPP can be used for this watershed, the effects of land-use practices can be precisely evaluated and alternative land management practices that will significantly improve water quality of Lake Decatur can be identified. In 1992, the City signed a Letter of Commitment (LOC) to the Illinois Environmental Protection Agency (IEPA) to achieve compliance with some of the water quality standards no later than April 15, 2001. Watershed management is identified as one of possible approaches listed in the LOC. We believe that if WEPP can be used for the entire watershed, proper watershed management practices can be identified and their effectiveness in reducing water quality and sedimentation problems can be accurately evaluated.

Time table and concrete results

Years 1-2: test DHARMA with WEPP on Lake Decatur watershed
Year 2: expand to SITES with USDA

4.4 Educational infrastructure

DHARMA will be a mechanism for innovative improvements both as a system to be explored, and as a tool for use in the classroom. Computer science students will see its design, and construct new features, while scientists use its functionality to learn the lessons of environmental research.

Computer scientists Students in Computer Science have already been heavily involved in the creation of the prototype as a real-life distributed system, and that involvement is expected to continue. We have developed the course *Advanced WWW Technologies* over the past few years. The course is designed to give students experience with the dramatic advances in Internet technology occurring daily. During the Spring 1999 semester more than ten graduate students spent much of the semester designing and developing the underlying prototype code. In the 1999-2000 school year, several seniors implementing the Java-based user interface to the system depicted in Figures 2. We plan to continue involving the students in system design and development, allowing them to get practical, hands-on experience while exploiting emerging technologies. The course is popular with many non-traditional distance-learning students, who are expected to bring the technologies and techniques learned from the system into the workplace.

Hydrological science educators We will evaluate various educational uses for our system through collaboration with local and statewide science educators. At the post-secondary level, we plan to improve future science teachers by allowing them to see research in action through on-going participation in research. By immersing them in a research experience, we believe that they will see science as a disciplined way of knowing, not just dry knowledge. Furthermore, this experience will carry forward into their own teaching.

USDA field engineers We will collaborate with the USDA to provide training for USDA engineers at regularly scheduled meetings.

Results and tools will be widely disseminated using the World Wide Web, and other means.

Time table and concrete results

Year 1: Develop functionality with CS and Hydro students, begin assembly of educational materials for field engineers.
Year 2: Continue with CS students; and hold workshop to train USDA engineers and scientists.

5 Conclusion

In the course of this project, we will develop a sophisticated distributed computation system which allows hydrological scientists and engineers to address fundamental questions utilizing distributed data without expending the prohibitively large resources required with currently available systems. The object-oriented framework will allow us to integrate different simulation models into a large scale system, and balance the computational load across the Internet. We will integrate several scientific models into a large-scale system.

This project will advance fundamental research in two areas: hydrological sciences, and computer and information sciences.

1. Hydrological and environmental science:
    - define a metadata schema for scientific data, in particular hydrologic data
    - provide new data and metadata to complement other research efforts
    - evaluate the effectiveness of the system as a research tool for exploring complex, distributed modeling environments
2. Computer and information science:
    - explore the issues of automatic data acquisition
    - design object-oriented framework for integrating watershed objects and simulation models into a large-scale system
    - design a distributed, run-time environment to schedule distributed data transmission and computational load across metacomputing systems and other resources

In summary, the end result of the project will be a distributed domain-specific middleware system, based on Internet standards, for improving the ability of environmental researchers to explore questions which are infeasible at the present time. The system will improve our understanding of the computing and information requirements for developing comprehensive, complex simulations in an object-oriented system. It has the potential to revolutionize hydrologic modeling on the watershed scale, leading to an improved environment for all. Additionally, the project will provide the foundation for involving a wide range of students, educators, and engineers in scientific research and discovery.