Strategic Management Plan for the

CUAHSI Hydrologic Information System

11 December 2006

1. Introduction

This plan presents a working strategy for continued development and management of the Hydrologic Information System (HIS) begun in 2003 as a collaborative Phase I award to the University of Texas (David Maidment, Principal Investigator) with support from the San Diego Supercomputing Center, Drexel University, the University of Illinois, and the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI). Phase I explored the many dimensions of emerging cyberinfrastructure and sensing technologies that need to be brought together to serve hydrologic science and related communities effectively in data discovery (screening the wide range of available information and planning how to integrate needed data into a common system), data delivery (entering measurements by others and data obtain by sensors at newly organized observatories into this system), and data publication (making data in the system available to users).

Phase II is awarded as a continuing grant for five years contingent on annual reviews to a team led by David Maidment at the University of Texas with subcontracts to the San Diego Supercomputing Center, Utah State University, Drexel University, and Duke University. This single continuing grant assigns management responsibility to Dr. Maidment to provide NSF a single point of contact for managing information system development. This award is jointly supported by the Hydrologic Sciences Program (L. Douglas James), the Environmental Engineering Program (Patrick Brezonik), and the EAR Information and Facilities Program (David Lambert); and all three programs will be participating in the reviews. The current grant to CUAHSI corresponds to the first two years of the Phase II HIS award, and continued CUAHSI funding will be requested in a renewal proposal.

CUAHSI has complimentary roles in identifying and expressing community needs to the HIS investigators and in disseminating information on HIS capabilities and distributing HIS products. CUAHSI is an organization of about 120 universities and research centers largely supported by the Hydrologic Science Program at the National Science Foundation to articulate needs and facilitate provision of community infrastructure and services. During Phase II, HIS will still largely be in a developmental stage; and the primary users will be faculty, researchers and graduate students testing system functionality at CUAHSI member institutions and other US universities with hydrologic science programs. Other users will be served as financial resources permit. CUAHSI does not receive funding through the HIS award but through a separate grant.

HIS activities will be coordinated with other NSF programs for environmental observations. The primary connection is with the WATer and Environmental Research Systems (WATERS) network, a joint effort of the Engineering and Geosciences Directorates to pursue capital funding through the Major Research Equipment and Facilities Construction (MREFC) account. Issues in HIS interfacing with field programs involving sensors and in supporting research in science and engineering are being explored at eleven funded observatory "test bed" sites that are exploring issues in. Additional awards are expected in 2007 to establish two Critical Zone Observatories that will provide testing in serving interdisciplinary research. The National Ecological Observing Network (NEON) and the Ocean Observing Initiative (OOI) are other major emerging initiatives using observations of freshwater quality and quantity to pursue research an education in related disciplines. Other NSF awards through other programs will also be connected over time to the HIS effort.

2. Cyberinfrastructure for Community Water Science

2.1 Mission

The purpose of the Phase I study was to outline a system of cyberinfrastructure to support Hydrologic Science. Development of the system and initial software to begin supporting operational capabilities will be pursued in Phase II. This second phase will draw on other experiences at NSF in developing cyberinfrastructure and such data centers as Unidata and IRIS that provide data conduits to academic institutions and tools for data interpretation. Considerable exploratory work is still needed to institutionalize an effective strategy for NSF to work with researchers to form cyberinfrastructure serving many users with diverse needs. The long-run vision is to transform science to be able to address water issues at the larger scale now needed to meet the major challenges that society faces in water resources planning and management. The short-run approach is to work with the WATERS network observatory testbed sites in designing a system that captures the wide and increasing variety of data that are being collected and to distribute those data to the community.

In contrast to other domain sciences where academic investigation is largely based on measurements by researchers, water science depends heavily on water measurement programs operated by public agencies at federal, state and local levels. Multiple federal, state and local agencies measure streamflow, water quality, groundwater, and weather and climate conditions. In studying any watershed or water body, data collected by individual scientists must be interpreted in the context of other data collection programs.

The information needs of water science are further complicated by the fact that fresh water exists in three very different environments. Measurements of water in the atmosphere, at the land surface, and in the subsurface, are accumulated in separate data systems designed to reflect the characteristics of these different environments. Atmospheric water moves rapidly in a four-dimensional space-time continuum over the earth; surface water moves more slowly in focused flow paths through watersheds, streams, rivers, and water bodies; and groundwater percolates very slowly through layered hydrogeologic formations containing macropores and fractures. Humans have added an extensive built system of reservoirs, conduits, and treatment plants to the surface water system. Thus, the measurements must cover many types of data in diverse environments.

Measurement design is further complicated by the need to capture how the properties of moving water are altered by characteristics of the surrounding environment. The physics of water movement and the chemical and biological processes that alter water quality vary considerably depending on local soil, geology, terrain, stream system and land cover.

In this context of partnering among multiple institutions that work in diverse physical, chemical, and biological environments, the mission of HIS is

> To build an access and sharing system for academic and public water observation data and data about the water environment and to enable the linking of data and models to understand how water systems function.

2.2 Partnerships

HIS must operate through multiple working partnerships. During Phase I, the HIS team built partnerships with government agencies that collect or use water data, with industry needed to develop the commercial software to support an operating system, with supercomputer centers that must design and provide needed web services and databases, and with domain sciences from many academic disciplines. These partnerships are illustrated in Figure 1, and those already underway are described below.

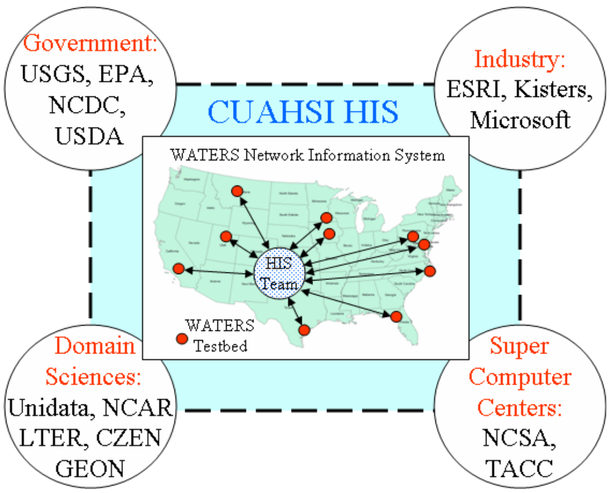

Figure 1. Illustration of CUAHSI HIS Partnerships

Federal Agencies. During Phase I, key partnerships were formed with the US Geological Survey's National Water Information System for water resources data (streamflow, water quality, groundwater); with the Storet (Storage and Retrieval) system in the EPA Office of Water, which stores water quality data; and with the National Climatic Data Center, the nation's repository for weather and climate data. Each of these agencies is collaborating with CUAHSI to provide web services and observations catalogs of their water measurement sites. The Agricultural Research Service of the US Department of Agriculture is designing a similar data system to HIS called STEWARDS to archive and publish the data from its experimental watersheds. Discussions are ongoing with USDA-ARS as to the best means of linking HIS and STEWARDS.

Commercial Software Firms. HIS must rely on commercial enterprises to prepare the software needed to support an operating information system that can service the large numbers of requests for data that will surely come. An initial partnership was made with the Environmental Systems Research Institute (ESRI), which is the manufacturer of ArcGIS, the geographic information system used within government agencies to produce geospatial datasets describing the water environment, such as the National Hydrography Dataset, National Elevation Dataset, National Land Cover Dataset, and a recent new dataset that combines all three of these, NHDPlus. HIS intends to use NHDPlus to form its digital description of the land base for hydrology. ESRI is donating its own resources to program the HIS Server Map Viewer. This map interface enables an integrated view over observation measurement networks operated by different agencies and individuals. Other commercial partnerships are emerging with Microsoft Corporation for storage of observational data in the SQL/Server relational database, with Kisters (a German firm that is the largest global company supplying software systems for water sensor data), and with the Danish Hydraulics Institute (a primary source of water modeling software). The expectation is that these commercial enterprises will be self financing until they become able to support themselves by selling HIS services to large numbers of users outside the CUAHSI community.

Supercomputing Centers. The San Diego Supercomputer Center (SDSC) continues to have the key role in conceptualizing the needed cyberinfrastructure and producing the descriptive information needed by the software developers to produce the products needed for system operations. HIS Server will be implemented at the national level at SDSC the San Diego Supercomputer Center and will also be deployed at the regional scale at each of the testbed sites. Another partnership is with the National Center for Supercomputer Applications at the University of Illinois, whose Environmental Cyberinfrastructure Demonstration project takes data from CUAHSI web services and shows how to analyze those through integrated applications. A third is with the Texas Advancement Computation Center, which has compiled in relational database format a complete history of the calibrated Stage III Nexrad radar rainfall, and is providing access to this as a public web service.

Domain Sciences. A wide variety of partnerships have already been established with the domain sciences, and many more will be added as HIS matures. These are described below.

Testbeds. The 11 WATERS network observatory testbed sites are pursuing science goals as they work with HIS in developing supporting cyberinfrastructure. The variety of needs that HIS will be planning to serve can be seen in the following themes:

- Testbeds at Baltimore, the Sante Fe watershed in Florida and the Susquehanna basin in Pennsylvania are concerned with surface water-groundwater interaction and need data systems that describe the subsurface environment;
- Testbeds at Iowa, Utah State and Minnesota are focused on real-time water quality monitoring in watersheds using sensor networks, and the Susquehanna testbed has a similar interest in sensor networks but more focused on water quantity;
- Testbeds at Montana, in the Sierra Nevada of California, and at Utah State will need descriptive data to cover snow depth, spatial coverage, and melt;
- Testbeds at Corpus Christi Bay, Texas, and the Albemarle-Pamlico Sound, North Carolina, are concerned with real-time and historical water quality measurement in large water bodies.

The HIS team is forming partnerships with the testbed teams to assess the model of using centers of expertise to provide specialized support. By following this model, HIS will become the hub of a WATERS Network Information System where much of the specialist knowledge and expertise is at the testbed nodes. Appendix A contains a memorandum describing the anticipated interaction of HIS and WATERS Network Information System, including the role of a Coordinating Committee intended to periodically review and report to NSF about this interaction.

In 2007, two additional NSF awards are expected to establish Critical Zone Observatories (CZO's) that will also become connected to the WATERS Network Information System. Other related projects or "ProtoObservatories" include the SAHRA Science and Technology Center at the University of Arizona (which has its own Hydrologic Information System activity supported by the Arizona Water Initiative); HydroKansas, a field site in Kansas with extensive hydrologic instrumentation; and a large research project at the University of Illinois that investigates water, energy and carbon fluxes at many sites across the nation. To the extent possible, HIS will also work with researchers at the Proto-Observatories to learn about their needs and share our technology.

The proto-observatories will introduce additional sorts of hydrologic data so that HIS can begin planning for the full complement of needed data. Important items include floodplain mappings, land use/cover, groundwater (depths, permeability, and pumping amounts in alluvium, fractured rock, and karst), vadose zone (soil moisture, soil characteristics, pedofunctions), geochemistry (nutrients, metals, organic chemicals), remote sensing, and built systems for water collection, storage, delivery and treatment. HIS will not, by any means, be capturing all these data types in Phase II, but it is necessary to be planning a system that can eventually achieve full coverage of the information needs for Hydrologic Science. HIS will use its connections to the proto-observatories to planning a system that can capture these widely divergent needs.

Complementary Systems: HIS is also working with Unidata, whose service to academic atmospheric science is similar to the HIS goal of supporting better access to hydrologic data. Unidata has defined a data model, netCDF, which uses a binary file format used to store multidimensional arrays, and HIS has adopted this format as its default data description for water processes operating continuously in space and time. A member of CUAHSI's HIS team serves on the Unidata Policy Committee, which guides Unidata's operation and interaction with its federal data providers.

Another emerging partnership is with the LTER network whose office in Albuquerque supports LTER sites with information services. HIS is also working with CZEN (the Critical Zone Exploratory Network) whose coverage of the nation's soils is being adopted into the HIS system and for sharing information about soils and soil water parameters. Also, HIS is partnering with the GEON project, to develop geographic data models for groundwater and hydrogeology.

International Partnerships: The European Commission Water Framework Directive has supported the Harmoni project which has similar goals to CUAHSI. The HarmoniRiB database system is under development with the goal to enable the HarmoniRIB Data Centre to receive, quality control, store and make available the representative basin data to be assembled by the project. The HarmoniRib database uses an object oriented data model for the storage of watershed information system that appears to have considerable flexibility in the representation of observations data. Another European Commission Water Framework Directive project is HarmoniIT. This has developed the OpenMI standard for the linking of time series models. OpenMI is a well documented open source protocol for how models should exchange time series information (http://www.openmi.org/). There is much potential benefit for both CUAHSI and Harmoni projects through collaboration and sharing of information on observations data and modeling efforts.

The Australian Commonwealth Scientific and Industrial Research Organization (CSIRO, www.csiro.au) has supported the Water Resources Observation Network (WRON) which has similar goals to CUAHSI HIS, although its primary focus is for water resources management instead of hydrologic science. WRON (http://wron.net.au/), which is still under development, consists of a national accounting of Australia's water resources through the deployment of a national water sensor network, a national water forecasting system, and a national water reporting system. The deployment of this system will require research into sensors and data integration, advanced modeling and forecasting systems that utilize distributed data and grid computing technologies, and tools for spatiotemporal reporting and visualization of large quantities of environmental data. Like the Harmoni project in Europe, the WRON effort overlaps significantly with the goals of the CUAHSI effort and both efforts will benefit from collaboration and sharing of experiences.

New Partnerships. Other partnerships are being formed with data projects or centers that serve particular functions. An ITR project at the University of Iowa provides data access and interpretation capabilities for Stage II Nexrad data (i.e., the radar reflectivity data and its conversion to rainfall estimates at each radar site). Other partnerships will be pursued to incorporate remote sensing data from NASA. Digitizing remote sensing data is currently very time consuming but needed by many domain sciences, and HIS will pursue meeting this common need through cooperative arrangements with other potential users.

Summary: All these partnerships are vital to the HIS mission. They supply in-kind resources and services that multiply by many times the direct investment that NSF makes in funding HIS.

2.3 The Dual Function of Cyberinfrastructure

HIS cyberinfrastructure must foster innovative research while serving users needing consistent information through time. Academic research in hydrologic science poses innovations, uses a plethora of methods for analysis and visualization, moves to the next concept, tries computer code and modeling in various programming languages, and improves the system step by step. Researchers in the hydrologic sciences also need dependable high-speed access to consistent and reliable data through time. This need dictates methodical migration from one software version to the next to ensure continued reliable sustainable operations. The HIS team supports these needs for *sustainability* and *flexibility* by building separate components: HIS Server and HIS Analyst, and linking them through web services, as shown in Figure 2.

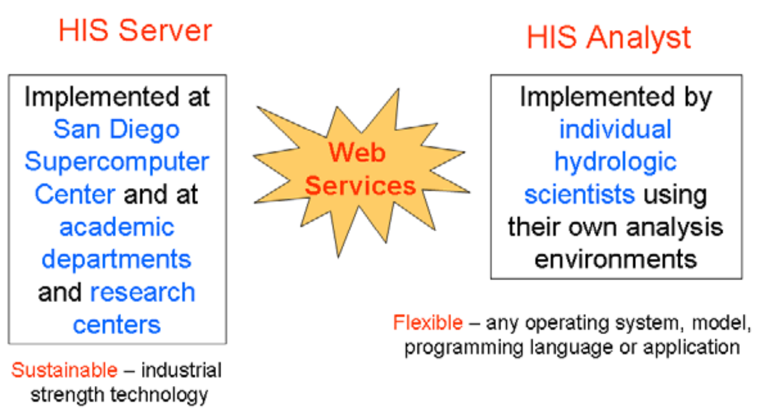

Figure 2. HIS Server and HIS Analyst are linked using web services.

HIS Server is built using industrial technologies (ArcGIS Server and Microsoft SQL/Server) and CUAHSI's web services technology, is implemented as a back-end data server at a research center or academic department, and is maintained by their computer system administrators with support from the HIS team. HIS Analyst is a set of templates and interfaces that researchers can use to apply CUAHSI web services in pursuing innovative ideas for system improvement. Because the HIS Server and Analyst components are linked through the internet using standard web services protocols, individual hydrologic scientists can use any operating system, hydrologic model, programming language or software application on any operating system that they choose. This preserves the *flexibility* of the scientist and the *sustainability* of a core server technology.

2.4 Future Challenges

Phase II: HIS will continue to build and develop web services and products, to develop and maintain the national HIS at SDSC, to support the HIS Server implementation at the testbed sites, and to develop and maintain HIS Analyst at Texas. The major challenge after Phase II is to develop the software needed to serve a large community of users that could eventually number many times the CUAHSI membership with an increasing variety of data.

Others: Phase II funding does not provide sufficient resources to undertake the following components of a fully functional HIS:

- *Build and maintain a support center* for HIS deployment at CUAHSI institutions that goes beyond servicing the testbeds. Separate provision will be needed for responding to phone calls, emails, and public wikis bringing questions from faculty, researchers and students at CUAHSI institutions or elsewhere.
- *Contract with commercial software firms* for providing on a sustained basis the products and services that ESRI and other firms are now donating to support the development of CUAHSI HIS. HIS needs the strong support of commercial software firms and a significant component of consultant services and commercial contracting will be needed to sustain HIS indefinitely.
- *Establish a public data server*. The wider water research and management community in the nation also needs the WaterOneFlow services through the data window that HIS Server provides. HIS and CUAHSI must work with our federal water data partners to provide public access to our services.
- *Presenting an Annual HIS Symposium*. As part of its Phase I activities, HIS organized two national HIS workshops in Austin, Texas, the first in March 2005 and the second in November 2006. These were both well-attended with about 80 participants from outside Austin, and some additional local participants. The funds to support travel and local expenses to these meetings were provided by the CUAHSI Program Office as part of its core grant. To continue this activity in future years will require either continuing the financial support through the CUAHSI Program Office or a separate grant to the HIS program.
- *Meetings of Review Committees*. It is important that HIS continue to interact with its federal data partners, and neighboring science cyberinfrastructure programs. Funding for meetings for this purpose has in the past been provided by the CUAHSI Program Office. This will either need to continue or be replaced by funding to the HIS program for this activity to continue.

These are serious needs that will require additional funds beginning in FY08 and increasing in FY09 and later. The current HIS budget is less than one third that provided to companion cyberinfrastructure projects such as BIRN, GEON, Unidata and NEES. A plan is needed for developing these services if HIS is to become an operational system.

3. Execution Plan for HIS Phase II

3.1 Project Definition

HIS is being developed to provide an informatics framework to advance hydrologic science by providing easier access to a more comprehensive data base (joining data collected by government agencies with different missions and by academic scientists from different disciplines), by providing data organization for seamless integration, and by providing tools for data analysis and visualization. Information science is required to advance process science. Hydrologic science will be advanced by supporting the representation of hydrologic processes with equations (process science) by an enhanced capacity to describe hydrologic environments quantitatively with data (information science).

3.1.1 Specific Goals

Phase I focused on a conceptual design that would establish web services to access to time-series data published by Federal agencies and to publish time-series data measured by university researchers. Phase II will expand on the initial conceptual design to develop specific designs that software developers can use to establish an operating system. The work pursues six specific goals:

a. Advance Hydrologic Informatics. The purpose driving HIS is to enhance hydrologic science by facilitating user access to data providing more reliable and more detailed representation of fluxes and environments for testing hypotheses and analyzing processes. For building this data base, Hydrologic Informatics links physical, chemical, and biological conceptualizations of the environment with measurements that can be used in focusing modeling on science issues. Hydrologic Informatics works through digital representations to gain understanding of hydrologic processes. In phase I, information science methods were developed for coupling fluxes across disparate scales and processes in a digital watershed, and for mining and analysis of information from the coupled system. In phase II, Hydrologic Informatics will develop digital representations of hydrologic features such as stream reaches, snow packs, estuaries, and groundwater plumes that can be used to integrate space and time representations in modeling.

b. Develop and Sustain Partnerships. As an organization of the academic community, CUAHSI relies strongly on the partnerships presented in Fig. 1 and Sect. 2.2. Much of the success of HIS can be attributed to forging partnerships, and Phase II will continue to strengthen the old and build new relationships. Although CUAHSI will lead these efforts, HIS resources will be used to support cooperative activities with the partners.

c. Develop HIS Server Functionality. Phase I focused on prototyping server capability for data delivery through web services and viewer-based data discovery. Phase II will add capability for data discovery through semantic/ontology based search and discovery web services, address data publication through the addition of web services to accept user data into the observations data model, and add functionalities for integrating information from HIS server nodes deployed at other locations and for serving spatial fields and space-time data. These functionalities will be advanced to release status.

d. Deliver Data Services from the National HIS Server. Phase I established preliminary capability at the national HIS Server maintained at SDSC that hosts the national WaterOneFlow web services: the national data catalog that indexes the data available and the Hydrologic Data Access System (HDAS) internet portal and map viewer. Phase II will stabilize these designs and advance the services to release status in the form of a national Observations Data Model (ODM) that will permit investigators to discover data sets, extract and download data subsets, and connect with HIS workgroup servers. Phase II will also add data sources, both from national networks and from observatories and WATERS testbeds. The added functionality will be delivered from the National HIS Server.

e. Deploy HIS Server at Proto-Observatories . A deployable version of HIS server, referred to as the HIS Workgroup Server will be deployed at WATERS test beds and possibly other proto-observatories. This will be a scaled down version of the national HIS server derived from the same code base. The workgroup server will serve local data, and a local workgroup catalog will index the subset of national data selected for the local observatory from national databases. A local ODM will host local data. HIS workgroup web services will provide access to national data in the local catalog as well as data from the local ODM. HIS workgroup will also include the data loader, ODM database tools, and other tools for Workgroup HIS administrators.

f. Develop HIS Analyst. The HIS Analyst is a suite of client software analogous to Microsoft office. HIS analyst provides individual users enhanced analytical capability through HIS web services for using existing software in a familiar analysis environment. HIS Analyst will primarily consist of a suite of tools and plug-ins for software such as ArcGIS, Matlab, Excel and R. For some systems, HIS Analyst functionality will be brokered through the HydroObjects library installed on the client as part of HIS Analyst.

3.1.2 Component Activities

These goals will be pursued through: (1) Research; (2) Technology Development; (3) Deployment and (4) Community partnership and outreach.

1. *Research.* Research has important roles in developing the needed Hydrologic Informatics, in sharing ideas and experiences with partners, and in generating new ideas and informatics concepts to build HIS Server and HIS Analyst functionality. HIS Analyst is expected to considerably enhance community research capability.

2. *Technology Development*. The technology development will add both server and analyst capability in pursuing the long-term goal of a comprehensive integrated Hydrologic Information System that supports discovery, analysis, integration, modeling and visualization of hydrologic data. Special efforts will be made to

capture data from remote repositories and from research projects. Specific activities are to establish:

- Common data formats for observation data, fields and geospatial themes
- Data retrieval, extraction and transformation enabling hydrologic analysis, modeling and visualization
- Cross data source search and discovery of observation data, fields and spatial themes
- Data sharing, archiving, and integration by using web services on HIS nodes.
- Ability to construct "digital watersheds" as integrated views over available local and remote data resources and services
- Ability to annotate data resources and services using common vocabularies

These activities will be pursued by SDSC working closely with the Environmental Systems Research Institute (ESRI) to develop open-source technology for HIS Servers. The fourfold goals are (1) to integrate HIS research findings into commercial software, (2) to utilize ESRI's software development powerhouse and market reach for the benefits of this project, (3) to streamline software support for the proto-observatories, and (4) to create a separable, modular and robust platform on which CUAHSI HIS developers can prototype and deploy additional research software.

3. *Deployment*. Goals 4, 5 and 6 are pursued by deploying HIS technology that captures intellectual advances in using hydrologic information. The deployments are at three levels:
   - HIS National Server at SDSC
   - HIS Workgroup Servers at the proto-observatory sites
   - HIS Analyst Desktop Applications Suite available to the community.

   Each deployment will be preceded by a stringent three-stage review process to ensure the reliability of the software. The review levels are:

   1. <u>Development</u>. These are services and products that are under development and have not gone through quality control checks. Only the HIS team and agency or observatory collaborators get to see and use these. The development server at SDSC with URL http://water.sdsc.edu/wateroneflowdev/ will be used to develop these web services and products.

   2. <u>Provisional</u>. These are products and services that are provisionally complete and have gone through a quality control check by the HIS team. They can be accessed and used by others on the understanding that they are still evolving. They may not be "feature complete" in that new capabilities may still be added, but the existing capabilities should work properly. They may be undergoing formal review prior to release. The provisional server at SDSC with URL http://water.sdsc.edu/wateroneflow/ will provide access.

   3. <u>Release</u>. These are services that are "feature complete" and have gone through a formal documented independent technical review. These products are

considered to be fully functional and will be accessed through the production HIS server at SDSC with URL http://river.sdsc.edu. They will also be available at the CUAHSI website http://www.cuahsi.org under Hydrologic Information Systems.

4. *Community Partnership and Outreach*. The shared nature of the water observation infrastructure makes partnering nationwide crucial. A bottleneck in hydrologic data delivery is found where WaterOneFlow GetValues service calls are routed to agency databases via web service wrappers at SDSC. This model necessitates re-tuning the services whenever an agency changes its web delivery system. The goal in partnering is to establish the CUAHSI HIS portal and services as a discovery interface with agency data while supporting agencies implement their own GetValues services. Our collaborations with USGS and EPA follow this scenario, while our NCDC collaborators have already begun developing their own WaterOneFlow services following our specification (http://www7.ncdc.noaa.gov/webservices/CUAHSIServices.html). Another component of the collaboration is establishing and maintaining agency observation data catalogs at the SDSC Enterprise HIS server. Discussions will continue with these three agencies on catalog maintenance and web services, and additional agencies will be brought into the discussions over time.

   A common issue in technology development is establishment and adoption of effective standards. Many database schemas exist for storing data at experimental watersheds. Many have custom tools that have been developed and repeated across sites. A goal of HIS is to promote the use of common data formats and protocols for interaction between data and models to maximize the opportunities for interoperability. We will examine alternative standards, develop methods consistent with established standards (OGC's GML, ISO, SI units, etc), and use these within HIS.

ESRI will continue through Phase II to devote its resources to develop server and client software capability serving the hydrology community and is fully collaborating with HIS to insure that software products meet user needs. Their role is very important to the long-run impact of the HIS investment by NSF.

3.2 Organizational Structure

HIS Phase I was a conceptual development and prototyping project involving a large team of academic domain PIs and cyber-scientists at SDSC. Phase II involves fewer domain scientists working with SDSC in producing usable information products. During the last two years, these team members worked together by communicating by email and bi-weekly conference calls. The past division of labor of having the academic PIs do the research and prototyping and the San Diego Supercomputer Center consolidating and synthesizing products will continue in Phase 2. As PI, David Maidment will retain direct fiscal control over the entire project. David Tarboton of Utah State University is Deputy Leader of the project. The other team members will participate through annual subcontracts awarded through the University of Texas at Austin. In August of each year, Dr Maidment will review all activities in the project and recommend to NSF a budget with subcontracted amounts to the partner institutions that reflects the judgment arrived at to that time as to the best utilization of NSF funds for the coming year. CUAHSI is not an awardee nor a subcontractor but will play a central role in prioritizing activities to meet community needs and in facilitating distribution of HIS products to the research community. Rick Hooper, the president and executive director of CUAHSI, will direct these CUAHSI activities, and will participate in the annual internal project review just described.

3.3 Advisory Structure

The HIS project will work with two associated committees, an *HIS Standing Committee* and an *NSF WATERS Network Information System Coordinating Committee*.

CUAHSI has a standing committee in Hydrologic Information Systems that will provide community input and oversight for HIS. This Committee has been functioning through Phase 1. Its members are Dennis Lettenmaier, University of Washington (Chairman); Larry Band, University of North Carolina; William Michener, University of New Mexico; Paul Morin, University of Minnesota; and Kelly Redmond, Desert Research Institute. When necessary, the Standing Committee develops formal reports to the CUAHSI Board about HIS progress.

The WATERS Network Information System Coordinating Committee (CC) oversees coordination between the proto-observatories and HIS and provides a venue for the PIs on these projects to give formal input on prioritization and conduct of HIS activities. Its members are Jean Bahr (University of Wisconsin), Peter Goodwin (University of Idaho), David Goodrich (USDA-ARS, Tucson), and Pedro Alvarado (Rice University). Detailed information on the interactive roles of HIS, CUAHSI, the proto-observatories, and the Coordinating Committee is attached in Appendix A, "The Waters Network Information System."

3.4 Roles and Responsibilities

The needs of Hydrologic Science are articulated by domain scientists at the University of Texas (UT), Duke University, and Utah State University (USU). The UT-Austin team, led by the project PI, is responsible for general coordination and the HIS Analyst suite of tools. Duke will work on methodology for using HIS resources in legacy and original hydrologic models. USU has primary responsibilities for data publication, the observation data model (ODM), and related user tools. The San Diego Supercomputer Center (SDSC) is primarily responsible for cyberinfrastructure development and deployment, technical coordination within and beyond the HIS project, and infrastructure integration with related cyberinfrastructure efforts. SDSC will collaborate with ESRI on HIS Server aspects of software development. Drexel University provides expertise in metadata, ontology and semantic mediation to support data searching and interpretation.

3.5 Work Tasks

3.5.1 Research in Hydrologic Informatics

This research will look for more efficient methods and models for discovering, locating and working with hydrologic information in pursuing and transforming research in hydrologic science. This research will be informed and prioritized based on interactions with the proto-observatories to identify informatics problems that need to be resolved to support observatory development. Some specific topics, with the assigned team within HIS in parentheses, are:

Data models for time-space data fusion in digital watersheds (All);

Representation of continental scale water processes and the data support required for implementation of hydrologic models of flow in streams for the entire continent (UT);

Cyberinfrastructure involved with linking environmental sensors to observations data models and creation of bi-directional feedback between sensors and HIS (USU);

Semantic Mediation and Global Data Search and Discovery (Drexel);

Controlled vocabularies and metadata for semantic annotation of hydrologic information (Drexel)

Data models for representing added hydrologic features in environmental observatories; comprehensive integration between hydrologic observations and geographic representations at the landscape scale, including context-based addressing to locate observations relative to their surroundings such as meters downstream from the junction of streams A and B, rather than just an x,y point in geographic space (USU);

Water quality modeling combining legacy models with Bayesian Networks and SPARROW (http://water.usgs.gov/nawqa/sparrow) concepts to utilize the HIS for management and organization of spatio-temporal data (Duke);

Modeling concepts developed specifically for HIS service-oriented architecture. This exercise will investigate a number of work-flow and model integration tools being produced for service-oriented modeling such as Kepler, OpenMI, CyberIntegrator, and ModelBuilder (Duke);

Integration with web services and portals maintained by related earth sciences projects, and design of service oriented architecture and computational framework supporting comprehensive hydrologic modeling within digital watersheds (SDSC).

3.5.2 Community collaboration to sustain, develop and pursue HIS partnerships

1. Participate and engage with partners both formally and informally to develop mutually beneficial agreements on standards for web services, access to and dissemination of data through CUAHSI WaterOneFlow.
2. Develop specifications for tool interfaces to HIS so that users may develop and contribute tools to the HIS Analyst toolkit and tools to interact with ODM
3. Develop policies for the free sharing of and access to technology developed by HIS
4. Develop procedures for community participation in and contribution to HIS through websites, wiki's etc.
5. Host an annual workshop to report progress and receive feedback from users, evaluators and partners. Users and evaluators will initially comprise the

environmental observatory test bed teams as well as others interested in implementing and evaluating HIS within their organization. The Navajo nation department of water resources will specifically evaluate the use of HIS for meeting Navajo nation needs. Commercial collaborators such as ESRI and Kisters as well as major federal data providers such as the USGS, EPA and NCDC will also be invited.

6. HIS will support for the growing community of hydrologic software developers by providing access to HIS systems for shared software development, documents, guidelines, and web service templates.
7. Maintain the websites for HDAS deployment as well as development, provisional and release stages of HIS system prototyping.

3.5.3 Development of HIS Server Functionality

The software architecture used to provide HIS server functionality is illustrated in Figure 3. The complete system that comprises web services, the web portal and map server interface, database servers and software to retrieve data from remote locations is referred to as the HIS Server. At the center of Figure 3 is the block depicting WaterOneFlow Web Services. Web services are computer applications that interact with and exchange information with other applications over the internet. The CUAHSI web services, named WaterOneFlow, are designed to be a standard mechanism for flow of hydrologic data between hydrologic data servers (databases) and users.

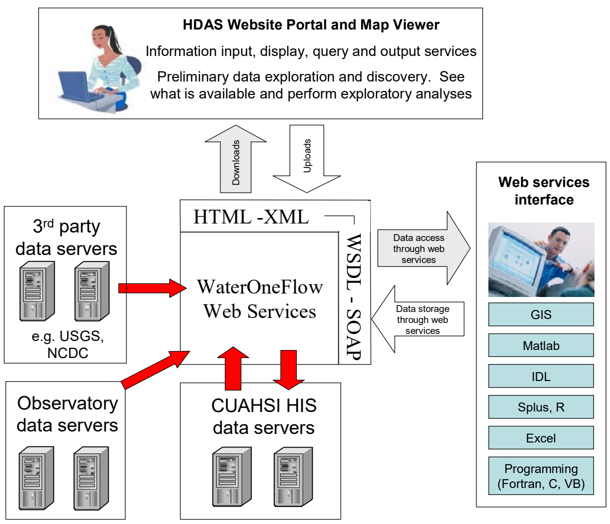

Figure 3. HIS Server data delivery architecture

Three forms of servers are illustrated: (1) third party servers for national databases, such as the USGS National Streamflow Information Program and National Climate Data

Center (NCDC); (2) Observatory servers holding data from hydrologic or environmental observatories; (3) CUAHSI Servers holding general collections of CUAHSI data. The block at the top of Figure 3 depicts the Hydrologic Data Access System (HDAS) Website Portal and Map Viewer. This is a website (located at http://river.sdsc.edu/HDAS/) accessed in the conventional way using an internet browser designed to facilitate the display, exploration, query and discovery of available data.

The block at the right depicts the web services interfaces providing access through a wide range of programming environments. They are designed to provide direct machine to machine access to capabilities available over the internet, and make data access more efficient by getting the browser out of the way. The WaterOneFlow web services are designed to make the capability of the CUAHSI HIS available from within the analysis or applications programming environment of a users choice, thus minimizing the additional learning required for user access. Web services provide cross platform capability for users with varying levels of programming sophistication.

Figure 3 also depicts, between the web services interface and HDAS, the communications protocols that facilitate the data transfers that occur. The website portal uses hypertext (HTML and XML). The web services use the simple object access protocol (SOAP) with web service capabilities described using the web services definition language (WSDL). The arrows in Figure 3 depict the flow of data. WaterOneFlow web services will ingest data from a range of data servers and then make them available for download from HDAS and to the web services. There will also be the capability for users (properly authenticated) to upload data using HDAS, or directly from web services. This data will be saved in the CUAHSI HIS data servers. CUAHSI HIS will not have upload capability for third party or independently operated observatory servers in Phase II, although third party servers may themselves support the same website portal and web service capability that the CUAHSI HIS does. The capability for HIS Server to integrate information from other data servers allows for the federation of data servers. Functionality will be developed for the national HIS server to provide federated integration of data across multiple HIS workgroup server nodes deployed at WATERS test beds and other proto-observatories.

Phase I developed the architecture for HIS Server and developed web services to the provisional level for the following national databases.

- USGS National Water Information System (Daily Values, Groundwater, Real time unit values and instantaneous irregular data)
- DAYMET daily surface meteorology from the National Center for Atmospheric Research
- MODIS - Moderate Resolution Imaging Spectroradiometer instrument aboard the Terra (EOS AM) and Aqua (EOS PM) satellites
- NAM - NCEP North American Mesoscale (NAM) Weather Research and Forecasting (WRF) Model.
- EPA Storet water quality data

CUAHSI HIS web services calling signatures have also been adopted by NCDC for their direct data dissemination. Some development web services are available for additional data sources and the CUAHSI observations data model.

The CUAHSI HDAS map viewer developed in Phase I has the capability to serve data from two data sources: (1) USGS National Water Information System (NWIS) daily values and (2) EPA Storet. The map viewer provides a map interface for initial exploration of the sites where data is available. Preliminary data exploration functionality is provided by the internet based NWIS time series analyst.

The HIS Server functionality task is divided into three components: (1) Observations data model; (2) Hydrologic Data Access System and Map Viewer Portal; and (3) WaterOneFlow Web Services. The tasks for each of these are described next.

Observations data model and tools

The observations data model (ODM) is a relational database schema designed to store hydrologic observations and sufficient ancillary information (metadata) about the data values to provide traceable heritage from raw measurements to usable information allowing them to be unambiguously interpreted and used. Phase I developed a prototype ODM schema. Phase II will undertake the following ODM tasks:

1. Evaluate revisions to ODM based on interactions with test bed project participants.
2. Develop ODM use cases from test bed participants
3. Develop additional ODM tools to support test beds, for example automated ingestion of real time data into the ODM
4. Develop tools for migrating database versions, and for migrating data between HIS nodes and archives
5. Develop sensor-to-database capability to support test beds. Coordinate with test bed groups to define requirements. Develop standards for describing sensors based on SensorML that will enable automatic discovery and integration of sensor data with the ODM.
6. Improve the capability and sophistication of the ODM data loader tool to improve data annotation (metadata), and metadata interpretation. ODM Data Loader is software for loading observation data from tab-delimited text, CSV (comma-separated value), and Microsoft Excel files, into an instance of the ODM.
7. Develop interactive tools for visualization, management, manipulation, and editing of data in the ODM
8. Integrate search wizard capability and ODM
9. Evaluate expansion of the data model to include flexible attribute definitions and flexible value types for observations
10. Evaluate generalization of data types to include vectors, rasters and other generalized data structures
11. Extend the ODM to provide more comprehensive integration between observations and the geographic representation of observatories and observation sites, including more context-based addressing.

Hydrologic Data Access System

The Hydrologic Data Access System (HDAS) is a web-based application providing discovery, extract and download access to multiple hydrologic repositories available via HIS server and data access web services. The HDAS will be available nationally from the enterprise version of HIS hosted at SDSC as well as deployable as part of the workgroup HIS to regional observatories and WATERS test beds. With the first version of HDAS prototyped during Phase I, our focus in Phase II is to:

1. Enhance HDAS in response to feedback from test bed teams.
2. Develop a formal specification for the HIS server catalog, listing observation networks and spatial data layers available through the server.
3. Develop capability for using queries to select data based on standard attributes in catalog schema, and a mechanism for saving the selections and materializing the selected data as downloadable database.
4. Develop tools for HIS server configuration.
5. Prototype and deploy communication between central and workgroup HIS servers; develop capability to integrate search and visualization based on data from remote nodes
6. Develop services for HIS workgroup node monitoring and management and deploy them at the HIS central server.
7. Develop capability for semantic mapping to support search of the catalog and ODM through development and use of hydrologic ontologies and tools for semantic annotation.
8. Prototype HIS integration with data publication and search components of other cyberinfrastructure projects, such as GEON and ROADNet

Software development in HDAS will be shared with ESRI. ESRI is developing the Map Viewer components of the Website Portal.

Web Services

Web services provide a standard and uniform interface to diverse data sources and types of data. The standard web service interface should comprise data discovery, data delivery, data query, data transformation and data upload functionality. The sequence of steps in web service development includes:

1. Formal definition of the calling and function return signatures for web services, for each of these functions.
2. Review of the web services definition document and prioritization of functions based on survey of potential users from test bed teams and other collaborator focus group.
3. If the service provides access to externally managed data (e.g. USGS NWIS, EPA STORET, etc.) – collaborate with agency personnel on web service specification development.
4. Prototyping of the web service against ODM and national data source databases
5. Development of supporting data catalogs and web service registration
6. Implementation of the web service in WaterOneFlow
7. Web service validation, testing and documentation

WaterOneFlow web services development will expand on the prototype services developed during Phase I of the project. The initial set of service signatures comprised

GetSiteInfo, GetVariableInfo, GetValues methods, as well as several convenience methods supporting the prototype version of HDAS. Table 1 gives the status of CUAHSI Web Services at the completion of HIS phase I. The following section provides more information on the tasks in web services development.

Table 1. Status of CUAHSI Web Services

	Delivery		Information (Metadata)						Publication			Discovery		
Data Source	Getvalues	GetValuesObject	GetSiteInfo	GetSiteInfoObject	GetSites	GetSitesXml	GetVariableInfo	GetVariableInfoObject	PutValues	PutVariableInfo	PutSiteInfo	GetSiteList	GetVariables	GetRecordsetWithSQL
USGS NWIS	P	P	P	P	P	P	P	P						
DAYMET	P	P					P	P						
MODIS	P	P					P	P						
NAM	P	P					P	P						
EPA Storet	P	P	P	P	P		P	P						
NCDC	D	D	D	D	D	D	D	D				D	D	
CUAHSI ODM	D		D		D		D						D	D

Development and testing status is indicated above by the letters following.

P. Provisional. Tested by HIS team and available for evaluation by outsiders on http://water.sdsc.edu/wateroneflow/

D. Development. Undergoing development and testing by the HIS team.

R. Release. Has passed review and released for general use (no services are at the release level)

The shaded boxes indicate web service and data set combinations that are not compatible so will not be implemented. Specifically we will not have publication services and record level query capability for third party datasets and do not provide site information for spatial fields not associated with specific sites.

Data Publication services for observation data

- Design method signatures, develop and test *PutValues*, *PutSiteInfo, PutVariableInfo* web service methods, for populating ODM and observation data catalogs from various sources, including catalog updates from federal agencies, data from instruments and sensors, researcher-supplied observation data, etc.
- Develop authorization methods for WaterOneFlow services, to enable query access to restricted databases and to support data publication

Data Discovery services for observation data

- Attribute-based discovery services, will utilize the semantic mediation work at Drexel and return variables associated with user-entered search terms, and stations where these variables are measured

- Location-based discovery services, returning lists of stations within a particular user-defined region (state, county, hydrologic unit, distance buffer of a linear or point feature, user-defined polygon)

Web services for other types of hydrologic and related data
- Develop or adopt web services for publishing and accessing collections of hydrologic vector layers, and incorporate them in HIS Server
- Develop or adopt web services for publishing and accessing climate fields, and incorporate them in HIS Server
- Develop or adopt web services for publishing and accessing remote-sensing data
- Develop or adopt web services for publishing and accessing data on features and performance of built environments including reservoirs, conduits, and waste treatment facilities.

Transformation services
- Develop web services for transformation of hydrologic vocabularies and units
- Enhance existing services with projection, units, time and vocabulary conversion capabilities

Additional tasks
- Explore standardization of WaterOneFlow services, in conjunction with OGC and OpenMI activities.
- Develop web services tuned to supporting additional HIS clients, such as Google Earth
- Provide support in adopting web services to specific needs of test bed projects

3.5.4 Deployment of National HIS Server

The informatics needs of the national hydrologic science community will be supported through the deployment of HIS Server National at SDSC at http://river.sdsc.edu. This server already has prototype status. Specific tasks in deployment of HIS server are

1. Testing and review of the prototype HIS Server to bring it to release status.
2. Deployment of new functionality developed under task 3.5.3 above and testing of this functionality through development and provisional to release status.
3. Acquire capability to serve additional data sources. The prioritization of data sources to serve will be based on surveys of user needs, both from the phase I needs survey and new surveys. More information on the likely scope and range of coverage is provided in the discussion on proto-observatories in Section 2.2. The steps involved are:
   3.1. Identify data source or network
   3.2. Collaborate with agency personnel on format and protocols for CUAHSI HIS access to the data
   3.3. Acquisition of site files and construction of catalog listing sites and variables
   3.4. Development of scraping or data acquisition methods in support of primary data delivery services (GetValues)

4. Enhance the National HIS Server with digital library, data publication, discovery and integration services being developed in related earth sciences projects (specifically, GEON)

3.5.5 Deployment of Workgroup HIS Server

HIS Server will be deployed at the WATERS testbed sites. Tasks are:

1. Finalize specification for server to be deployed at test bed sites in conjunction with test bed partners
2. Package and release initial HIS server software (version 1.0). Functional updates will be released once per year, with incremental updates to fix bugs.
3. Develop a specification of digital watershed as a collection of integrated views over several types of hydrologic data sources available in HIS server
4. Develop HIS server data catalogs for the specific regions comprising each deployment
5. Training of users and test bed data managers in the use of HIS Server.

3.5.6 Development of HIS Analyst

HIS Analyst is the client software that allows individual users to exploit HIS functionality and interact with HIS servers both national and workgroup. HIS Analyst is comprised of the following components:

1. HydroObjects. A dynamic software library for working from a PC to access through web services time series in a standard object format and providing transform functionality for reconciliation of units. A key functionality of HydroObjects is to provide capability for Excel to access web services. As additional web service functionality is added, HydroObjects will be developed to support this functionality.
2. Weather Downloader. This is a plug-in for ArcGIS to facilitate the downloading of Weather and Streamflow data from CUAHSI web services into ArcHydro time series data tables.
3. MyDB and MySelect are simplified table schemas. MyDB is a single table extracted from ODM to represent data values in a simple form suitable for analysis. MySelect is a table of sites and observation types from ODM, consolidated into a single table to serve as a basis for selection of the data an analyst wants to work with and in a format that can be returned to a web service to retrieve the selected data.
4. ODM and DataLoader tools. These tools will provide capability for interacting with the ODM through simple graphical interfaces similar to the Time Series Analyst developed during phase I (http://water.usu.edu/nwisanalyst/Default.aspx).
5. HIS Workbook. This is a tutorial style workbook that provides guidance on the use of WaterOneFlow web services from a number of analysis software and programming environments. The present workbook demonstrates capability for Excel, ArcGIS, Matlab, Java and Visual Basic. These will be extended to other web service enabled analysis environments deemed to be of high priority such as SAS, Fortran, C/C++, and R.

HIS tasks are

1. Investigate developing a more integrated HIS Analyst that provides the scientist a workspace for analyzing hydrologic data, either a new platform or using an existing extensible platform (e.g. ArcMap, MapWindow).

2. Develop integrated client based modeling capability using HIS functionality and model integration and workflow environments such as OpenMI, Cyberingegrator, Modelbuilder, Kepler.
3. Develop and evaluate procedures for common data reduction tasks, such as interpolation, regular gridding of irregular data, handling missing data. This might lead to a "recipes for working with real messy data" publication and tools as part of HIS analyst that implement this.

3.6 Broader Impacts

HIS will have a profound impact by facilitating the networking of hydrologic scientists at many universities contributing and receiving hydrologic information. HIS data and tools will be developed in the public domain and be made available to the professional hydrologists and educators at all levels. HIS also demonstrates how cyberinfrastructure can serve related earth and environmental sciences. Already, related NSF Environmental Observatory initiatives (OOI/ORION, NEON, and LTER) are closely following the cyberinfrastructure HIS developments. Moreover, the goal of uniting the nation’s water information has wide value for water management, engineering and planning, and HIS helps the data delivery services of all Federal water agencies reach greater potential. HIS also partners with the Navajo Nation water resources program to integrate and publish their water information. In the computer industry, CUAHSI HIS has earned attention as an example of building distributed information systems with scientific principles that supports natural resources management.

3.7 Reporting

An annual report will be provided to NSF and the oversight committees. An abbreviated version will be published on the CUAHSI HIS website. HIS will continue to maintain the internal document and code sharing space begun in Phase I. It will provide minutes of conference calls, discussion documents, papers, presentations, and group reports.

NSF will arrange site visits involving the HIS team, NSF Program Officers, and a selected site-visit team towards the ends of the second and of the fourth year of this award. The site visit at the end of the second year will involve the WATERS Network Information System Coordinating Committee and will evaluate accomplishments (based in part on data supplied relative to the metrics listed below), achievements through interaction with the proto-observatories, needs to be potentially addressed through a second round of observatories, and plans for HIS during the rest of the Phase II award. The send visit at the end of the fourth year will evaluate accomplishments to that point and provide a forum for discussion on what should be done after the Phase II funding is exhausted.

3.8 Evaluation and Metrics of Success

Number of web service hits, by origin (geographic origin, com/edu domains, etc.)
Number of users of HIS workgroup services
Number of downloads of HIS software
Number of hydrologic software developers accessing CUAHSI code repository

Agencies adopting CUAHSI web services specification, or working with CUAHSI on hydrologic web services
Number and volume of datasets available through HIS
Number of presentations, training workshops, seminars

Year 1. Implement the HIS Server software at 10 testbed sites.

Quarter 1 (Dec 06 – Feb 07):
Test the initial version of HIS server for scalability, finalize server specification to be deployed at test bed sites
Complete data compilation and software development for the initial version of the HIS Server
Quarter 2 (Mar 07 – May 07): Initially deploy the HIS Workgroup Server at 10 testbed sites
Quarter 3 (Jun 07-Aug 07):
Complete deployment of the HIS Central Server at SDSC, with functional national observation catalogs and web services, and the ability to navigate and explore HIS Workgroup servers.
Quarter 4 (Aug 07-Nov 07):
Complete deployment of HIS Workgroup Server at the 10 testbed sites, with functional local observations databases, web services and observations catalogs for data sources in each testbed area

Year 2. Implement Digital Watershed assembly and analysis capabilities in the 10 testbed HIS workgroup servers.

The testbed-focused activities during this year will be structured similarly to year one. Specific tasks include:
Quarter 1 (Dec 07-Feb 08):
- design and formalization of digital watersheds as integrated views defined over available hydrologic resources for the area

Quarter 2 (Mar 08-May 08):
- develop methodology and tools for publishing additional types of hydrologic data via HIS workgroup servers (including field data and spatial themes)
- develop methodology for annotating the published data resources

Quarter 3 (Jun 08 – Aug 08):
- implement search over the additional types of hydrologic data, and extend the data cart implementation to include thus selected data
- implement the ability to export data carts as "digital watershed" databases that can be used as input for hydrologic analysis, visualization and modeling at watershed level

Quarter 4 (Sep 08 – Nov 08):
- test and deploy the digital watershed functionality in HIS workgroup servers at the 10 testbed sites

Beyond years 1 and 2, the CUAHSI HIS project will support further needs of testbeds by updating workgroup HIS Server software, developing additional software modules, and providing the ability to search and integrate with other sources of earth science data made available via parallel cyberinfrastructure efforts in geology, ecology, atmospheric sciences and other disciplines.

Glossary of Terms

Term	Definition
CUAHSI	Consortium of Universities for the Advancement of Hydrologic Science, Inc.
HIS	Hydrologic Information System
Web Services	Self-contained and modular web applications that interact with and exchange information with other applications, and can be invoked over the network using a standardized XML messaging system.
WaterOneFlow	The name for CUAHSI web services to support the standard mechanism for flow of hydrologic data between hydrologic data repositories (databases) and users.
Observations Data Model (ODM)	A general purpose schema and relational database used to store point hydrologic observations
HTML	HyperText Markup Language. The language used to create hypertext documents for the World Wide Web. In HTML, a block of text can be surrounded with tags that indicate how it should appear (for example, in bold face or italics). Also, in HTML a word, a block of text, or an image can be linked to another file on the Web.
XML	XML (Extensible Markup Language) is a standard for creating markup languages which describe the structure of data. It is not a fixed set of elements like HTML, but rather, it is a language for describing languages. XML enables authors to define their own tags. XML is a formal specification of the World Wide Web Consortium.
SOAP	Simple Object Access Protocol. SOAP is a standard for exchanging XML-based messages over a computer network, normally using HTTP. SOAP provides a basic messaging framework used by web services.
HTTP	HyperText Transfer Protocol. A protocol used to transmit files over the World Wide Web.
WSDL	Web Services Description Language, an XML-formatted language used to describe a Web service's capabilities as collections of communication endpoints capable of exchanging messages.
HIS Server	Complete system for access of data supported by CUAHSI HIS, comprising data servers, data catalogs, WaterOneFlow web services, website portal and map viewer interface.
National HIS Server	Implementation of HIS Server at SDSC on national databases
Workgroup HIS Server	Deployment of HIS Server at an observatory, test bed or research group
HDAS	Hydrologic Data Access System. Website portal and map viewer interface component of HIS Server
Development	Development stage where products (software or web services) are undergoing testing by the HIS team and not yet suitable for sharing with the community
Provisional	Development stage where products (software or web services) have been developed and tested within the HIS team and are suitable for a wider range of testing by others within the CUAHSI community; but their performance and reliability have not been fully evaluated
Release	Development stage where products (software or web services) have been tested both within and outside the HIS team and their reliability and performance are assessed as being suitable for distribution to the wider community.
Data Loader	Software for loading data into the observations database
ODM Tools	Software tools for the manipulation of data in the observations database
Catalog	The schema and database that holds ancillary information necessary to index sites and data series available using HIS web services
HIS Workbook	Document that provides tutorials on the use of HIS
HydroObjects	Dynamic software library for a PC to simplify the use of web services from some analysis software, notably Excel

APPENDIX A
THE WATERS NETWORK INFORMATION SYSTEM

Overview:
The NSF programs in Hydrologic Science and Environmental Engineering are joining in making 11 awards to support "Test Beds" to evaluate ideas for effective deployment of sensors to enable research at the watershed scale. Simultaneously, these programs are making an award for continued development of the Hydrologic Information System (HIS) to provide a common portal serving researchers throughout the United States working at the watershed scale (whether now at the Test Beds, later at environmental observatories, or on other projects or at other sites). The concept behind the portal is to provide these researchers easy access to data from the Federal agencies and other miscellaneous sources and to provide a forum that they can use to exchange data and information amongst themselves. In addition, NSF is supporting efforts by the hydrology (CUAHSI) and the environmental engineering communities (CLEANER) for planning the WATERS Network observatory system. Other programs at NSF are supporting observatory development for the ecology community (NEON), for Ocean Sciences (OOI), and for Critical Zone Observatories (CZO). In this context, it is important to have a working relationship in which HIS and the Test Beds work together toward designing the WATERS Network and at the same time exchange information within this larger context for the common benefit. Implementation details were further discussed at the Test-Bed kick-off meeting in Austin on November 15-17, 2006.

The purpose of the Test Bed awards is to address technical and logistical issues for cost-effective observatory design. The purposes of the award to HIS are (1) to address issues in providing the cyber-infrastructure needed for storing and communicating information across the observatories and (2) to obtain context data from agencies that researchers can access through a single point of contact. The peer review of the proposals for Test Beds selected proposals to test important issues in defined geographic areas. At each location, cost effective testing requires inventorying and accumulation of data that are relevant to formation of the WATERS Network and that others have already collected for the area. A database with that information will serve the local Test Bed team, the other Test Bed teams, other researchers, and resource managers.

The database for a Test Bed is to include (1) data obtained by HIS from the federal agencies (USGS, NOAA, EPA, etc.) so as to save the Test Bed teams considerable time and effort, (2) data that the Test Bed team finds from local sources (largely state and local agencies, power companies, and past researchers), and (3) data that the Test Bed team collects with its installed sensors. HIS will obtain data of the first type to begin building the data base for the geographical area specified for each Test Bed. Each Test Bed is expected to enter its data and supporting meta data in the second and third categories into "data files." The goal here is not to add to the burden by requiring the Test Beds to go out and find additional information; it is to provide storage for the information obtained. The criterion as to which information should be collected is one of relevancy to the issues being examined at the Test Bed based on the concept that a Test Bed should not be setting out sensors to obtain information that is already available, directly or indirectly, from other sources.

The database that grows at a site is what we are calling a "digital watershed," and HIS will be responsible for Digital Watershed management. Digital watersheds will support the tests run by each Test-Bed team as well as studies by other researchers and practitioners working in the region or making nationwide comparisons. Over time, the approaches to collecting, describing, and accessing data will evolve to serve users better. We will all benefit from (1) access to context information for Test Bed watershed-based field studies, (2) information for nationwide comparisons of ideas in different contexts, (3) a resource that will facilitate collaboration in multi-disciplinary research and (4) an information source for educators and water resources managers. This final application is important for enhancing broader impacts.

The framework for a digital watershed will be established by the Hydrologic Information System (HIS) to support ready exchange of data among Test Beds and widespread access by others. In addition, HIS will work with Test-Bed PIs to determine and accommodate site-specific needs. Ideas that start locally will become valuable nationwide.

While we have a vision of making a digital watershed into a comprehensive source of information for real watersheds, we recognize that a truly comprehensive database is far beyond the scope of a Test Bed project. Nonetheless, a start toward this vision will begin a useful legacy for future projects.

Testing Criteria:
The criteria to be used in evaluating ideas at a Test-Bed or for HIS is one of added value in the form of new hypotheses (conveying better understanding of the system) that can be tested or from reduced uncertainty in the outcomes of previously tested hypotheses. Some testing will involve comparing observations across several watersheds to reach general conclusions. A Coordinating Committee (CC) is to be formed by NSF to facilitate cross-test bed standardization of data entry (including metadata support) and of testing protocols. The CC will also facilitate discussion and peer review when differences in results among test beds seem problematic. It will suggest testing across Test Beds to gain insights that would not otherwise be gained.

NSF, with logistical support through the CUAHSI and CLEANER project offices, will have general oversight of the process and conduct annual reviews of progress by both HIS and the Test-Bed projects. At the end of year one, the annual review for a test bed will evaluate the work to date and Test Bed plans for activities for the following year. The annual review will also consider the performance of the system as a whole. Toward the end of year two, the review will address needs and opportunities for continuing funding.

Parties involved:

1. Hydrologic Information System (HIS), with David Maidment, U. Texas, Austin as PI.

2. Digital Watershed Management Team (DWMT) – Initially, the PIs (or a designated CI specialist) for the 11 Test-Bed projects. Later, other projects such as two or three CZOs may be added.
3. Coordinating Committee (CC) – A committee to be appointed by NSF after considering recommendations from CUAHSI and CLEANER.
4. Programs in Hydrologic Science and Environmental Engineering at the National Science Foundation (NSF).

HIS Functions in Supporting Digital Watersheds:

1. *Data Discovery, National Level* – Facilitate user access to the measurements and associated metadata from major water agencies operating nationwide (USGS, NWIS, EPA STORET, NCDC, and NRCS). Other agencies may be added later as resources permit. HIS will define each type of data included, the time period covered by data type (generally be the time period for which the agency has data in electronic form), and a map that shows areas covered by data type. This national database will include areas outside of the digital watersheds and work toward coverage of the entire United States.
2. *Data Publication, Watershed Level* – HIS will support entry (with first priority given to spatially-identified data from DWMT, second priority to spatially-identified data from other locations, and third priority to data from laboratory experiments that are not site specific) of data and associated meta data. The spatially-identified data may be obtained from others (Federal, state, local, or private sources) or be directly measured by the DWMT or some other university-based research project. HIS will provide necessary software. A DWMT may also derive relationships for use in developing secondary data (items not measured directly). These relationships and "calculated data" will also be entered along with metadata on statistical testing of the relationships. Also, the DWMT are encouraged to obtain data measured by remote sensing from NASA satellites or from aircraft. HIS will work with the DWMT in facilitating entry of this information into the system.
3. *Data Delivery* – HIS will provide data from "1" and from "2" into selected analytic environments (e.g., Excel, MatLab, SAS, etc.) through web services. Policies will be established with respect to charges for data (perhaps free to academic users and to agencies that have supplied data for HIS with a charge schedule for others) and to the kinds of help provided to users seeking access or interpretations for applying the information.

Interactions among the HIS, DWMT and CC:

Priorities for data entry are to be based on contribution to research supporting the WATERS network science plan or to the science goals stated in the proposal funded to establish the digital watershed.

HIS will:

1. Provide a standardized procedure for compiling data and metadata from outside sources and work with the DWMTs in problem troubleshooting.
2. Maintain the central information system.

3. Service requests for data delivery.

The DWMT will:

1. Advise HIS as to any problems related to the "national level data" for its area.
2. Provide to HIS available data at the watershed level.
3. Use data within the system (whether from its area or elsewhere) to test hypotheses in hydrologic science. Where a hypothesis cannot be tested, a DWMT will assess what additional information is needed, evaluate possibilities for obtaining needed data from the water agencies or others, or pursue a research project to obtain measurements not adequately covered.

The CC will:

1. Identify and reconcile differences in format, style, etc. among the DWMTs.
2. Review the total content of the data in HIS from the three sources and:
   a. Recommend research opportunities based on the contained information for two or more DWMTs.
   b. Recommend additional data that should be obtained from the nationwide water agencies
   c. Recommend changes needed to the metadata.
3. Suggest expansions to the DWMT programs (in either data gathering or research coverage) or to coverage by the national network.
4. Develop a program of information dissemination to encourage use of the information system and facilitate cross-DWMT scientific inquiry.

Conclusion:
The overriding program goal is for parties to join together to address the greatest and most difficult challenge for the WATERS network to transforming "doing science:"

** From having each investigator gathers data that he or she needs independently [often doing a poor job (ineffective parameter selections and weak quality control) on data types and information interpretations where they are less familiar] and present findings at conferences and through the literature;

** To having investigators use common resources (through a portal that gives and receives) where they exchange data interactively. The data base would be filled by data collected by experts in the field of study, and data items would have been exposed to a QA/QC program with a degree of scrutiny described in the metadata. Each investigator will have access to more data, more reliable data, and expert interpretations.

The central purpose in funding the Test Beds and in funding HIS is to facilitate everyone working toward this goal.