CONSORTIUM OF UNIVERSITIES FOR THE ADVANCEMENT OF HYDROLOGIC SCIENCE, INC.

TECHNICAL REPORT #2

CUAHSI
universities allied for water research

CUAHSI HYDROLOGIC INFORMATION SYSTEMS

Prepared by the CUAHSI Hydrologic Information Systems Committee

AUGUST 2002

CUAHSI HYDROLOGIC INFORMATION SYSTEMS COMMITTEE

David Maidment (Chair), University of Texas
Wendy Graham, University of Florida
Upmanu Lall, Columbia University
Anton Kruger, University of Iowa
Praveen Kumar, University of Illinois
Venkat Lakshmi, University of South Carolina
Dennis Lettenmaier, University of Washington
Michael Piasecki, Drexel University
Chunmiao Zheng, University of Alabama

Additional contributions were made by Jay Alameda, Marshall Moss, Hatim Sharif, Jeff Weber, and William Wise.

TABLE OF CONTENTS

Introduction and Rationale 1
Program Components 3
  Hydrologic Data Access Center 4
  Hydrologic Information Technology Program 8
  Hydrologic Information Science Center 14
Resource Requirements 18
Summary 28

INTRODUCTION AND RATIONALE

The goal of the CUAHSI Hydrologic Information System (HIS) initiative is to support innovation and advancement in the hydrologic sciences by providing better access to the hydrologic data and information systems technology needed to formulate and test new hydrologic science research hypotheses. This goal will be met at the level of the CUAHSI hydrologic science plan and at the level of individual investigators who are the core of advancement of hydrologic science. Synergy and feedback between these two levels is necessary to ensure the effectiveness of the CUAHSI Science Plan in the advancement of hydrologic science.

The CUAHSI Science Plan envisions advances in five key areas:

- *Critical processes at primary interfaces in the hydrologic cycle*, including land surface-atmospheric interactions, recharge and discharge of groundwater, and surface transport of sediment, nutrients, and toxic substances.
- *Behavior of hydrologic processes at different spatial scales,* especially at interfaces between hydrologic processes where the characteristic length and temporal scales of processes change abruptly.
- *Interaction between biotic and hydrologic processes* – the domain of a new science called ecohydrology.
- *Predictability of hydrologic systems* – the ability to use past and present hydrologic information to predict the future behavior of hydrologic systems.
- *Management of water resources* – the use of improved hydrologic science knowledge to increase the soundness of water policy and management decisions.

The CUAHSI Committee on Hydrologic Information Systems conducted a survey of the need for hydrologic data, information systems infrastructure, and services among the participants at the CUAHSI regional meetings. The results of this survey reveal a great diversity among academic hydrologists in the manner in which they use hydrologic information in their research. Nonetheless, some common desires are apparent from the survey. Among these are (1) better access to a large volume and variety of high-quality hydrologic data through the Internet; (2) access to visualization tools and data-analysis software to inspect and assess data; and (3) standardized data sets for hydrologic stores and fluxes across the United States that can be used as benchmark data sets for both individual and community model development. The complete survey responses can be viewed on the CUAHSI HIS website (http://www.iihr.uiowa.edu/~cuahsi/his/)

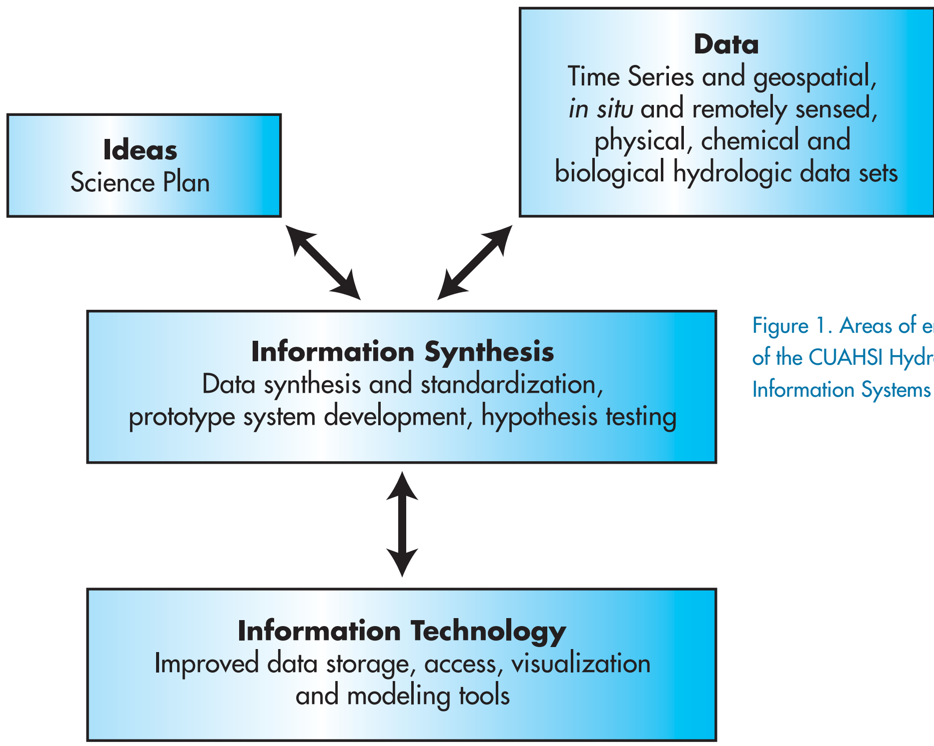


Figure 1. Areas of emphasis of the CUAHSI Hydrologic Information Systems Program.

One way to think about the mission of CUAHSI HIS is illustrated in Figure 1, which presents the CUAHSI Information Systems program as having four areas of emphasis: ideas, data, information technology, and information synthesis. The ideas area of emphasis comprises the intellectual insights, mathematical formulations, and motivations for research that animate the CUAHSI Science Plan and the work of individual hydrologic science investigators. The data area of emphasis encompasses all the bodies of data that are needed to conduct a particular investigation, whether obtained at the CUAHSI Hydrologic Observatories, by individual experiments, by hydrologic monitoring programs, or by remote sensing. The information technology area of emphasis encompasses the tools that are needed to assimilate, interpret, analyze, and visualize the data. The information synthesis area of emphasis is where the ideas, data, and technology are brought together to achieve new insights and scientific advances.

The task of the CUAHSI HIS Committee is to define an infrastructure to provide these services to the academic hydrologic science community to foster the application of information systems in hydrologic research to address new broad hydrologic science questions. This report summarizes the committee's progress to date and its proposal for the development of a CUAHSI HIS program.

PROGRAM COMPONENTS

The proposed Hydrologic Information Systems program comprises three operational components intended to address the data, information technology, and information synthesis areas of emphasis described in Figure 1. These components are:

- A Hydrologic Data Access Center (HDAC) that facilitates access to hydrologic data and provides information tools for data analysis. This component is seen as a physical facility located at or near a major university research institution containing computers and staffing to support hydrologic scientists throughout the nation, mainly by using Internet delivery of hydrologic data.
- A Hydrologic Information Technology Program (HITP) that designs and constructs high-quality hydrologic information system tools and integrated data sets for the use of hydrologic scientists. The tools and data sets will be distributed and supported by the HDAC. This component may be implemented by staff at the HDAC; by contracting with universities, government agencies, or commercial firms; or by a combination of these methods.
- A Hydrologic Information Science Center (HISC), that serves as a meeting place and collaborative research environment for hydrologic scientists to develop new information system prototypes, to define new data standards and methods for data set construction, and to interact with scientists in fields such as computer science and atmospheric science. This component is also envisaged as being a physical facility located at or near a major university research institution. It may be either located with the HDAC or separate from the HDAC. Because integrative science is at the heart of CUAHSI and not just of its Hydrologic Information Systems component, this component may become part of a similarly conceived center of larger scope supporting CUAHSI as a whole.

The interaction among these components is illustrated in Figure 2. Hydrologic scientists are directly supported by the Hydrologic Data Access Center. The Hydrologic Information Science Center would foster collaborative research as well as the development of new information system prototypes that are two to three years beyond what is currently available. The Hydrologic Information Technology Program would work with the Hydrologic Information Science Center prototypes and requirements from the Hydrologic Data Access Center to produce solid information tools and community data sets that are distributed to hydrologic scientists through the Hydrologic Data Access Center.

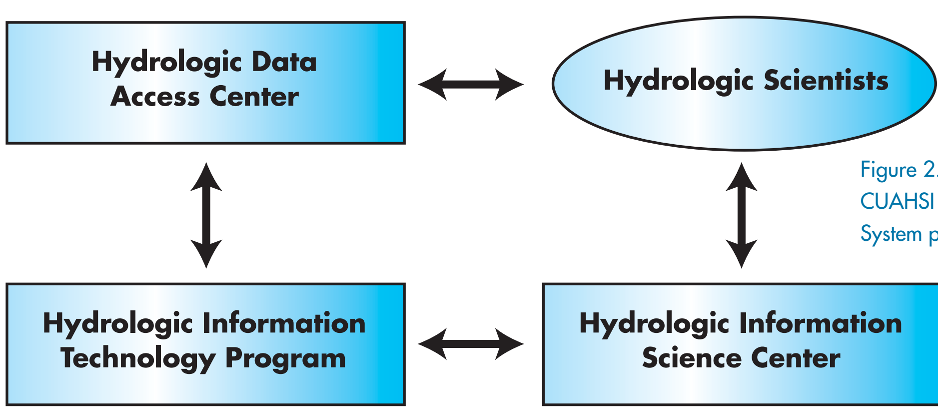


Figure 2. Interaction among the CUAHSI Hydrologic Information System program components.

HYDROLOGIC DATA ACCESS CENTER

The Need

Access to real time, systematic, *in situ*, distributed, proxy, and remotely sensed data for a full range of Earth system processes has been identified as a vital community need. This includes access to both existing data currently archived at universities and state and federal agencies, as well as new data that will be generated by CUAHSI-initiated programs. Access to existing data will require a methodology to reach out to data centers to query their data inventories, while the access to new CUAHSI-generated data will require development of an internal data repository that will store data specific to the hydrologic science community. To simplify access to multiple data sources, we aim to provide the capability to present the hydrology researcher with a single virtual data source.

CUAHSI's HDAC internal data system will need to address a number of issues at the core of data management systems. First, the center will have to identify the data products it seeks to provide to the user community beyond what can be retrieved from existing sources. Second, it will need to establish a standard for the description and storage/archiving system of these new added-value data sets so they can be queried, selected, and downloaded by external users. Third, these data need to be stored in such a way that retrieval is fast and comes in user-friendly formats. This is vitally important, as the success of the HDAC will depend heavily on acceptance by the user community that in turn will be shaped by basic performance parameters such as access time, ease-of-use, and scope, readability, and uniformity of data sets. Fourth, the HDAC will need to ensure quality control for both data received from other servers as well its own newly added data sets. Finally, it will have to provide a suite of online tools that allow inspection and analysis of the selected data prior to download and/or storage.

Technical and logistical support to the user community for accessing hydrologic data is a vital service of the HDAC and will include support for basic functions such as query and download options as well as support for analysis tool selection and data quality assurance. In addition, it will permit the user community to actively participate in the generation of new data sets to help expand the knowledge base of the HDAC data server. Active participation and inclusion of all users is an important concept that has shown its success in the open source and public domain community.

Experience in large-scale field experiments and research laboratories, for example, land-atmosphere field campaigns like FIFE (First International Satellite Land Surface Climatology Project Field Experiment) and BOREAS (Boreal Ecosystem-Atmosphere Study), and facilities such the Long-Term Ecological Research (LTER) sites, has shown that it is imperative to have an organized entity responsible for setting standards for data quality control, archiving, cataloging, and access. The proposed HDAC will be the entity charged with these responsibilities for CUAHSI activities.

Relation to the Science Plan

The underlying theme of the science plan is its emphasis on integration: integration across the interfaces among hydrologic processes, integration across spatial scales, integration between hydrologic and biologic processes, integration of past and present information to support prediction, and integration of scientific knowledge with water management and policy decisions. To achieve such a high level of integration requires that individual scientists and groups of scientists be able to link their information with that from neighboring disciplines or hydrologic environments, and thus form a more coherent picture of hydrologic process behavior than any individual or group can achieve alone. Information linking requires a comprehensive hydrologic information infrastructure that indexes data sources, provides rapid access to those sources, conveys data in readily usable formats, allows for the publication of new data sets generated through scientific research to be added to the common information infrastructure, and supports links to hydrologic modeling systems for simulation of process behavior. The Hydrologic Data Access Center will provide this service for CUAHSI.

How the Need will be Met

The Hydrologic Data Access Center will be a physical facility located at or near a major research institution; staffed with computer specialists, scientists, and administrative staff; and supported by a significant bank of data servers and Internet interfaces. The center will have a director and a governance structure to ensure responsiveness to the needs of the hydrologic science community and appropriate interfaces to government agencies that provide data needed by CUAHSI. The center will negotiate group contracts with software and data providers so that CUAHSI members can obtain low-cost or free access to software and data that otherwise would cost much more to obtain directly.

Program Elements

The program of the Hydrologic Data Access Center will include the following elements:

1. Assessment of User Requirements. This is a continuous process for interaction with users to ensure that the services provided are meeting their needs. The activity will be initiated with a more formal survey than the one conducted to date, and then later with a Committee of Users that meets periodically to guide the development of the Hydrologic Data Access Center.
2. Data Identification and Interpretation. This is a service that identifies what data is available on a global scale and where it is housed. Due to the heterogeneous nature of the available data sets, their description, querying, retrieval, and exchange requires the development of a metadata standard for hydrologic data. We would implement this as series of software components with web service interfaces so that we could provide a single software interface to the hydrologist, independent of the heterogeneous nature of the data sources feeding the service. The software components will also be able to transform the data into the format desired by the end user. Search, inspection, and visualization services will be delivered through a comprehensive web portal.
3. Data Archiving, Query, and Provision. In this service, the HDAC stores core hydrologic research data sets not archived elsewhere, including data from individual-investigator experiments and from the CUAHSI Hydrologic Observatories. These data will be archived on an HDAC internal hydrologic database and documented according to CUAHSI metadata standards. These data will be indexed alongside the other data archives described in element 2, so that with a single query the hydrologist can seamlessly query multiple data archives. Access will be provided to the data archive through the same web portal as envisioned in element 2. An example metadata database archiving, query, and provision architecture is seen in Box 1.
4. Hydrologic Internet Data Distribution. This function would parallel for hydrologic scientists the Internet Data Distribution services provided by Unidata to atmospheric science researchers (http://www.unidata.ucar.edu/). The system will provide an automated feed of diverse products from raingage networks and streamflow gages, via the Internet, to a subscriber's local machine. Where feasible, the system will automatically convert the data to the format of a user's choice, for example, plain text, netCDF, or HDF. In many cases the data will be available in real-time, as data become available from the sensors or providers. This may include data from CUAHSI long-term observatories and field experiments. In other cases, there will be latencies determined by quality-control measures and arrangements with data providers. The HDAC will create and maintain the core of the data distribution system. The system will be governed by a users' committee to make sure that day-to-day operations are meeting user needs. A separate policy committee will determine access to the system and act as a liaison with federal agencies, local governments, and commercial companies for data access, redistribution of products and data, and other matters. An overview of data distribution system implementation is shown in Box 2.
5. Training Programs. These training courses and other instructional materials to teach researchers and graduate students about available information sources and tools for processing the data available from those sources.

Box 1. Spatially Distributed Data and Computing Resources

For a number of years, computer scientists and engineers have been developing mechanisms to provide for secure communication, coordination, and access among a variety of computational and data resources. Many challenges exist as the heterogeneity of data descriptions, data storage, access protocols, and organization of computational resources have traditionally been carried out by those unaware of the need or possibility of making their node part of an overarching grid within which free access and interchange of information could be achieved.

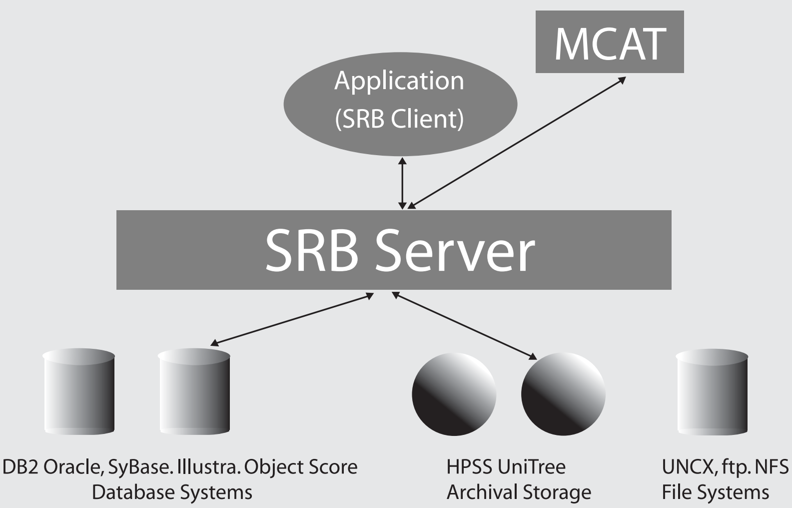
An excellent example of these new initiatives is the Globus Project (http://www.globus.org), which provides the basic infrastructure to distribute computational power among spatially distributed sites. Analogous to the electrical power grid, the computational grid is a combination of computing and data resources, linked through a common software infrastructure—the grid middleware—to act in a coordinated fashion. This project, the Alliance Grid Science Portal, is a prototype grid-based working environment focused on a higher level of abstraction of the computational grid to create reusable mechanisms for programming the grid from a computational scientist's perspective. The portal has the following features:

- It is accessible to both experts and non-experts
- It allows access to simulation codes, experimental data, and advanced tools including remote computers and distributed Linux clusters
- It desires to address substantial applications through collaborative efforts directed from the desktop

Another excellent example for activities that seek to unify and standardize data technology is the Scalable Information Networks Environment (SINE) workshop hosted by the Partnership for Biodiversity Informatics (http://www.sdsc.edu/pbi/index.html). A core element of the future needs envisaged by SINE is the development of tools to access highly heterogeneous data repositories for querying and retrieval. Implementation of a data grid service that seeks to ensure interoperability must address issues such as

- distributed storage access across heterogeneous devices
- data transport
- access protocols
- distributed computing services

At the core of this system is the Storage Resource Broker (SRB), a tool that acts as a virtual file system so the user can see where the data is stored and how it can be accessed. In addition, it provides mechanisms through which users can add their own data to relational databases without needing detailed knowledge of database structures.



CLIENT SERVICE MIDDLEWARE. The Storage Resource Broker (SRB) was developed at the San Diego Supercomputer Center (SDSC) and the National Partnership for Advanced Computational Infrastructure (NPACI) as client-server middleware to provide a uniform interface for connecting to heterogeneous data resources over a network and accessing replicated data sets. SRB, in conjunction with the Metadata Catalog (MCAT), provides a way to access data sets and resources based on their attributes rather than their names or physical locations.

HYDROLOGIC INFORMATION TECHNOLOGY PROGRAM

The Need

The CUAHSI Hydrologic Information Technology Program (HITP) will identify and develop new areas of information technology to advance the state of the art of hydrologic information science. Examples include data assimilation, data scaling, data mining, knowledge discovery, and geoinformatics. These are cross-cutting activities that will involve computer scientists, hydrologists, and geoscientists.

Development and maintenance of community-level predictive models has also been identified as a need by the hydrology community. Prediction lies at the heart of advances in hydrologic sciences—and all natural sciences. As hydrologists have moved toward a broader Earth system science view, surface water hydrologists have begun to recognize, for instance, that their prediction problems cannot realistically be isolated from those of groundwater hydrologists. Likewise, many of the cutting-edge problems in hydrologic and related sciences like fluvial geomorphology, ecohydrology, and others require the use of complex models that incorporate predictive capabilities in multiple subdisciplines of hydrology models (e.g., atmosphere, snow, canopy, land surface, vadose zone, and saturated zone). There is a need for development of common model interfaces and modeling tools that facilitate such cross-disciplinary research by allowing hydrologists in various sub-specialties access to state-of-the-art modeling tools outside their immediate areas of specialization. CUAHSI HITP will play a leading role in the development of such a common interface and linkage by organizing community-wide involvement and by publishing standards and protocols for model input/output structures and data exchange formats.

Essential to the efforts of both individual and community model development is availability of high-quality benchmark data sets. The use of benchmark data sets is well established in other areas of the Earth sciences. For instance, in the coupled land-atmosphere community, surface moisture, and energy flux data sets collected during field experiments like FIFE and HAPEX have been widely used for model testing. CUAHSI HITP is in a unique position to facilitate the development and maintenance of such benchmark data sets. Activities will likely include development of data formats and procedures for submitting benchmark data sets, peer review and quality control of submitted benchmark data sets, and maintenance of a website to host all available benchmark data sets. Development of modern self-documenting data archival formats like NetCDF and XML has facilitated exchange and access to large data sets in other areas of the Earth sciences, and could help promote hydrologic model development and testing.

Relation to the Science Plan

The accomplishment of the science plan requires tools for the intelligent investigation and manipulation of hydrologic information to support scientific inquiry and establish new insights. For example, at the interface between surface and groundwater, tools are needed to interface surface water data, which change rapidly in time and are associated with stream segments and land surface units, with groundwater data that change slowly over time and are associated with soil and hydrogeologic units whose spatial boundaries are quite different than those of surface watersheds. Understanding the impact of the hydrologic regime on aquatic ecology at a particular point on a stream network requires the estimation of a large set of streamflow statistics and stream channel conditions, and the capacity to understand how these statistics and conditions may change in response to land use change or manipulation of the hydrologic regime with different water management policies. Generalizing these ecohydrology results from individual points on a stream network to the whole stream system requires the capacity to integrate hydrologic information through space and time.

Understanding the behavior of processes at different spatial and temporal scales requires first that we have the capacity to transform data from one spatial and temporal scale to another; however, no generalized toolset presently exists to accomplish this task for hydrology. This task is complicated by the fact that hydrology has both a vertical water exchange component among atmospheric, land surface, and subsurface water, and a horizontal water exchange component in which water flows through the landscape of streams, rivers, lakes, wetlands, and aquifers to coastal water systems. There are many sets of overlapping geospatial boundaries for these hydrologic units, and geographic information systems are needed to be able to trace water movement appropriately from one unit to another. As shown in Figure 3, a hydrologic information system must simultaneously be able to deal with geospatial and temporal hydrologic data to describe the behavior of natural water systems, and it must have the capacity to provide linkages to hydrologic models or model components that will simulate process behavior and allow statistical testing of alternative hypotheses as to how this behavior is occurring.

Hydrologic prediction has been strongly supported during the past decade by linkage with atmospheric science. Numerical weather prediction relies on combining general circulation models of the atmosphere with data assimilation: the capacity for absorption of large volumes of current weather data, and adjustment of the states of the general circulation model before a new set of forecast model runs are made. Reanalysis projects have been conducted with the major general circulation models in which the current version of the model is run with weather data recorded during past decades to reconstruct a more comprehensive picture of atmospheric and land surface conditions. The great investment in land surface-atmospheric interaction models during the past decade has created a much more comprehensive picture of the vertical water balance than previously existed. This advance needs to be more comprehensively linked to the traditional domains of surface water and groundwater hydrology.

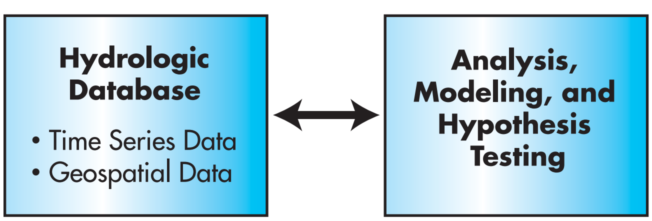


Figure 3. A hydrologic information system integrates geospatial and temporal hydrologic data and supports analysis, modeling, and hypothesis testing of process behavior.

Hydrologic research is most often focused on study of recorded data from past events, while atmospheric science is focused more on present and future information. A new set of tools and methods is needed in hydrology to complete the continuum of past, present, and future information. These tools must also take into account that the spatial area represented by a particular data entity may simply be a point (e.g., a rain gage reading), a line (e.g., a stream cross section), an area (e.g., a soil unit in a soil map), or a distributed measurement over a continuous surface (e.g., a Nexrad radar rainfall map). As shown in Figure 4, hydrologic information covers a continuum in space and time.

A traditional way that improved hydrologic science has been integrated with water management and policy formation is through its inclusion in hydrologic simulation models, such as the HEC-HMS and HEC-RAS models from the Hydrologic Engineering Center of the U.S. Army Corps of Engineers, EPA water quality simulation models such as HSPF and SWMM, USGS groundwater models such as Modflow, and USDA agricultural water models such as SWAT. Commercial software vendors such as the Danish Hydraulic Institute are now providing sophisticated hydrologic simulation software in the United States. Comprehensive new hydrologic information systems that combine data archives with querying and assessment tools, and a set of simulation models, such as the EPA BASINS system for supporting Total Maximum Daily Load analysis, which is built on top of the ArcView geographic information system are being produced. The hydrology profession relies heavily on having standardized hydrologic simulation models for which support and training are available.

Mechanisms are needed through which CUAHSI hydrologic information systems can be linked with standard hydrologic simulation models, and also for the creation of a new

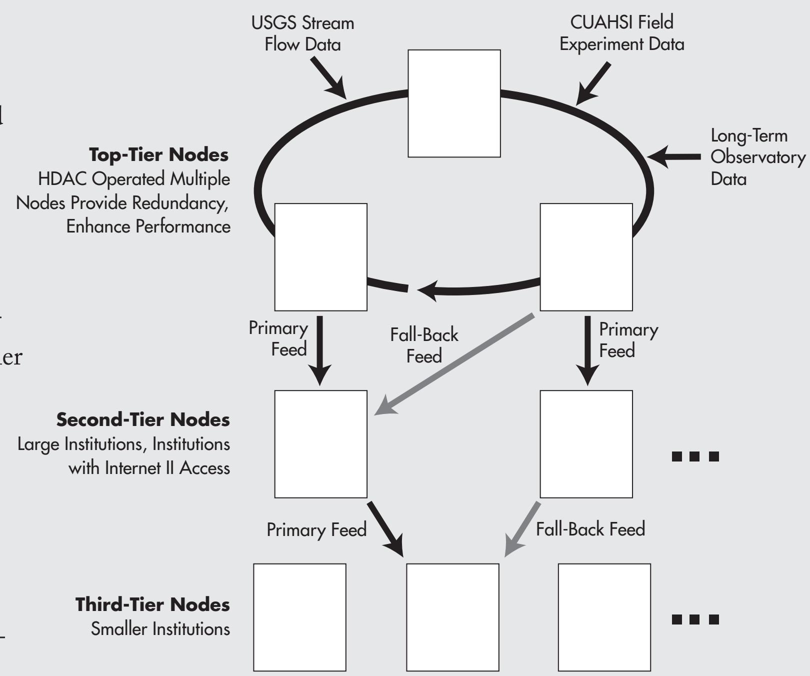
Box 2. Hydrologic Internet Data Distribution

Thr Hydrologic Internet Data Distribution system continuously conveys to hydrologic scientists a data stream describing the current hydrologic condition of the nation. This hydrologic condition is described in part by observational records, such as the USGS Water Watch system for real-time streamflow information, and in part by computed model states, such as soil moisture conditions derived by land surface-atmospheric interaction models. The data stream will be provided to academic and research institutions in a hierarchical fashion so that higher-level users receive a larger volume of data and lower-level users receive only a few selected data sets.

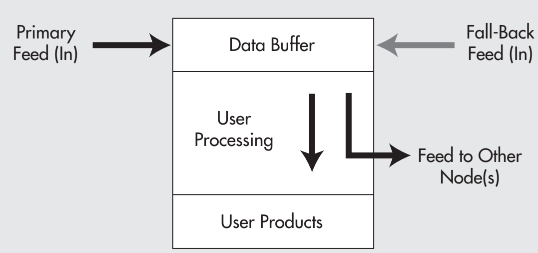
A top-tier ring is the primary recipient of the hydrologic data feeds. Multiple top-tier sites provide redundancy and improve performance. The HDAC is responsible for maintaining the top tier, which feeds second-tier nodes, which feed third-tier nodes, and so on. Nodes not in the top tier have at least one fail over upstream feed in case their primary feed fails. Upper tiers will handle more traffic so larger institutions are logical candidates, while smaller groups of users would be third-tier users. Users with modest data needs will be fourth-tier users.

Local Data Manager (LDM) software at nodes ingests data, strips out streams of interest, and feeds downstream users. The LDM buffers several hours/days worth of data as a safety net for downstream users in case of system outages. The system requires management by the HDAC personnel. The LDM also manages the data on the user's machine/node, and can organize data into a desired structure such as a hierarchical tree: site/year/month/data file. It can run external software to process the data into products and is bi-directional; with simple tools users can insert their own products into the data stream. An institution can thus process data and distribute the products on campus. HDAC will archive many, but not all, data it distributes. For example, it will not archive raw NEXRAD data as the National Climate Data Center archives the full data set.

The system is flexible, allowing easy addition of new nodes/users. It is inherently reliable due to built-in redundancy and follows a cooperative model that translates into low cost for individual users. All that is required is a low cost PC and the LDM software provided by the HDAC.



Internet Data Distribution Topology



Each Node Runs Local Data Manager (LDM) Software That Buffers, Strips, and Processes Data

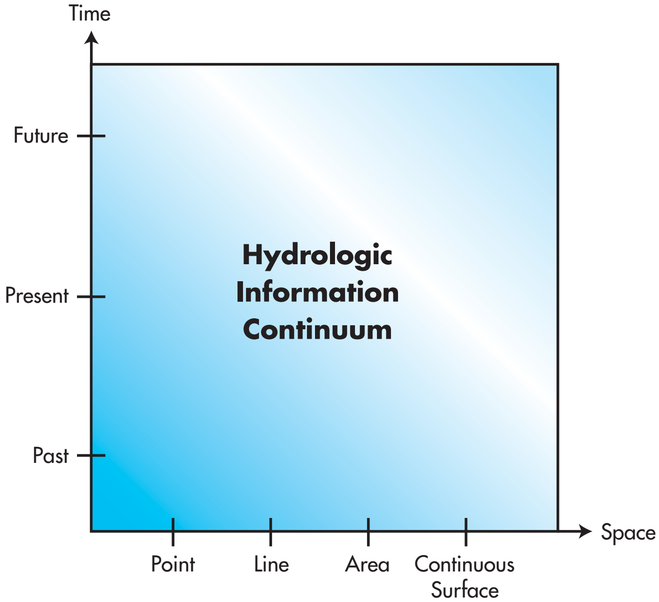


Figure 4. Continuum of hydrologic information in space and time.

CUAHSI community hydrologic modeling system in which component models created by CUAHSI investigators can be linked to an information infrastructure so that the results of one component can be used as the input data for another.

How the Need will be Met

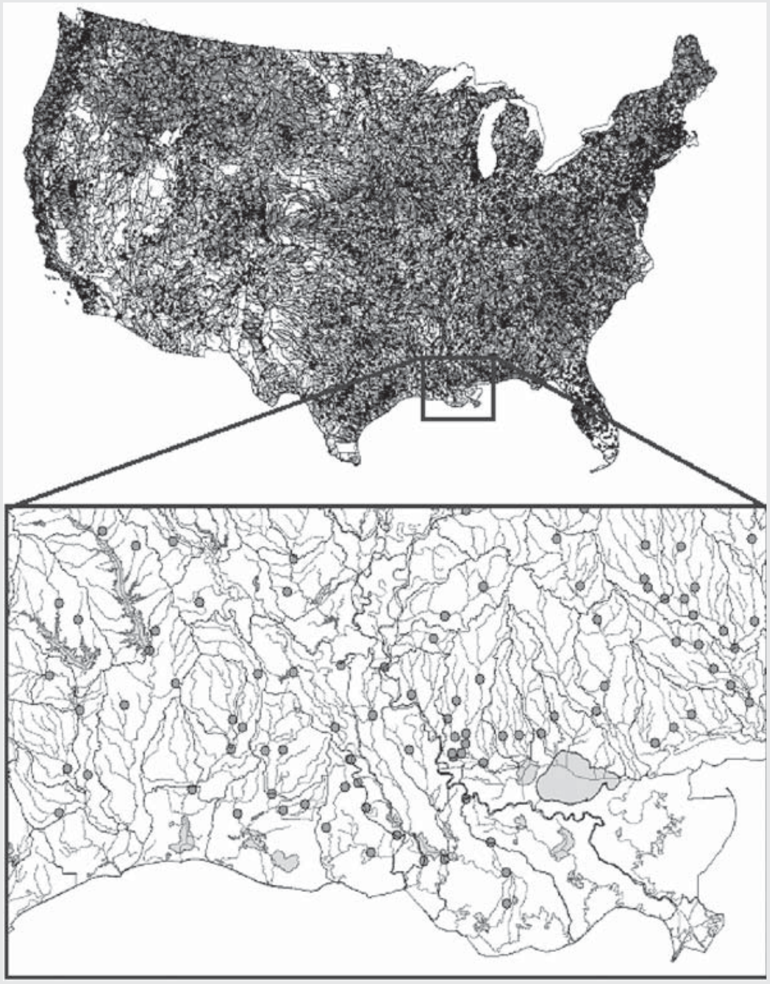
This need will be met by the development of a *Hydrologic Information Technology Program*, which will consist of a series of software components and data sets that are constructed to serve the needs of hydrologic scientists. The tools and data sets will be distributed and supported by the Hydrologic Data Access Center. The staff of the Hydrologic Data Access Center may undertake elements of this program, or they will be developed by contracting with universities, government agencies, or commercial firms, or by a combination of these methods.

Proposed Activities

- Hydrologic Information Toolkit. This is a software system specifically developed for CUAHSI HIS that would help hydrologic researchers automatically obtain structured data sets through URL connections, with tools for data conversion from one format to another. This would include a space-time toolkit to allow transformation of data from one scale of space and time to another, including conversion of data on flows between phases of the hydrologic cycle. Tools will be included for data assimilation, synthesis, mining, and visualization.
- Data Model for Hydrologic Science. This is a database design for storing geospatial and temporal hydrologic information from all phases of the hydrologic cycle in an integrated way. This database design will support development of tools for transforming hydrologic information among different geospatial and temporal frameworks. It will permit tracing of the movement of water and transported substances, and for the construction of mass balances among the phases of the hydrologic cycle. (see Box 3 for more information on hydrologic data models).
- Hydrologic Data Assimilation System. Data assimilation refers to various methods of merging sparse, intermittent measurements with models. It is now a mature field in numerical weather prediction and, more recently, has also been used by oceanographers. In the atmospheric and oceanographic science communities, data assimilation has come to be viewed as a scientific tool that can not only lead to better predictions, but can also help to diagnose model weaknesses. However, the realization that the synthesis of sparse data via models could yield similar benefits in hydrologic science is not yet widely acknowledged. We believe that one of the reasons for the slow development of data assimilation in hydrology is the difficulty that individuals and small groups of investigators have in accessing and handling large data sets. Therefore, we propose support for provision of access to real-time and near real-time surface and other hydrologic data sets, models suitable for production of assimilated fields of hydrologic fluxes and state variables, and for testing of new data assimilation methods specific to hydrologic applications. (see Box 4 for more information on data assimilation).
- Community Hydrology Modeling System. This is a mechanism for a creating community hydrology model with plug-compatible components to CUAHSI HIS data structures (See Box 5 for more information).

Box 3. Hydrologic Data Models

Data sets from many sources are needed to describe hydrologic processes and systems. A hydrologic data model is a structure or framework for arranging these data sets in such a manner that they can be accessed in an integrated way rather than as a series of separate data layers. Historically, hydrologic data have been held in binary data file formats custom-designed to support particular data sets or hydrologic model applications. More recently, large data archives, such as the USGS National Water Information System, have been converted into relational database management systems. Likewise, geographic information systems, such as ArcGIS, have adopted standard relational database management systems such as Microsoft Access and Oracle to store their data. Storing data in a relational database management system requires a system design in which the various types of data are stored in database tables, and the tables are linked using relationships, which are associations between identifying numbers for the data. Thus, for example, a GIS coverage of USGS streamflow gaging stations is related to a table of time-series data recorded at those stations by storing the USGS Site Number as an attribute of both the GIS coverage and the time-series table.



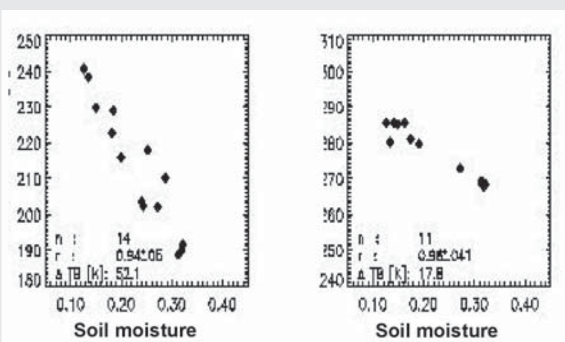
Rivers, waterbodies, watersheds, and streamflow gaging stations in Arc Hydro USA. (http://www.crwr.utexas.edu/gis/gishydro02/GISHydro2002.htm)

During the period 1999-2002, the Consortium for GIS in Water Resources (http://www.crwr.utexas.edu/giswr) formed by academic, agency, and industry partners, developed a database design for water resources within the framework of the ArcGIS geographic information system. This Arc Hydro data model supports both geospatial and temporal water resources data in a single integrated structure that can be applied at any spatial scale. As an example of Arc Hydro, a prototype application for the continental United States called Arc Hydro USA has been prepared that contains a topologically connected river network of the nation integrated with the principal lakes and waterbodies, 2000 Hydrologic Unit Code watersheds linked to their outlet points on the river network, 18,000 USGS stream gage locations (all active and inactive sites), and daily streamflow time series for 1527 streamflow gaging stations in the Hydro Climatic Data Network, each time series being related to the gage location at which it was measured. Integrated information resources like Arc Hydro USA make possible a level of inquiry into hydrologic processes in time and space across the nation that has not previously been attainable.

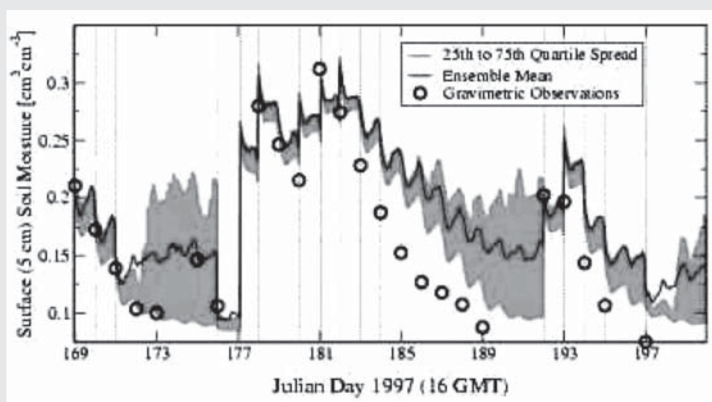
Box 4. Hydrologic Data Assimilation

Data assimilation refers broadly to methods that incorporate observations into a (generally deterministic) dynamic model to produce improved predictions or nowcasts at the time of assimilation, and subsequent improved (future) forecasts. It is widely used in weather forecasting to integrate disparate observations at irregular spatial and temporal intervals to produce better model initial conditions, and hence improved forecast accuracy. It is also being used increasingly in oceanographic and climate models. Data asssimilation's use in land surface modeling is relatively new, but it shows promise for improving hydrologic predictions by incorporating soil moisture, snow, surface temperature, and other remotely sensed and *in situ* observations. Data assimilation has also proven valuable as a diagnostic tool for evaluation of model error sources and for evaluation of optimal observation strategies.

Progress in hydrologic data assimilation (potentially applicable to soil moisture and temperature, groundwater, snowpack, groundwater, and streamflow prediction) has lagged behind its atmosphere and ocean counterparts. Some progress has already been made in improving hydrologic predictions using data assimilation methods, as illustrated in the figure. One potential benefit of data assimilation—unrealized as yet in hydrology—is its value in diagnostic studies, and subsequent improvement of model physical representations. Global and regional reanalysis is a powerful tool that has been used by the atmospheric sciences community, but it has not yet been exploited in hydrology, even though several well-known reanalyses do estimate land surface variables over global grids. One possible area where CUASHI could help lead future developments in the hydrologic sciences is to encourage the development of reanalysis products more relevant to land surface processes, and to foster their use to improve hydrologic prediction tools.



Sensitivity of microwave brightness temperature to soil moisture (SGP'97)



Soil Moisture Data Assimilation: Passive microwave sensors at relatively long wavelengths (e.g., L-band at 1.4 GHz) offer the best potential for observing soil moisture over large areas. However, satellite or aircraft platforms imply intermittent observations, and therefore the necessity to combine observations with models to produce fields that are continuous in space and time as required for hydrologic (e.g., flood forecasting) and meteorological (e.g., weather and climate forecasting) applications. The laft-hand panels illustrate the sensitivity to soil moisture in the SGP'97 experiment (central Oklahoma) of passive microwave brightness temperatures from L-band (high sensitivity) and C-band (much lower sensitivity). The right-hand panel shows results from a data assimilation procedure using the TOPLATS land surface model that ingests the L-band ESTAR (aircraft-based L-band sensor) microwave brightness temperatures. The drift of the model relative to *in situ* observations between ESTAR overflights (overflight times indicated by vertical lines in plot) is apparent. See Crow et al (2001) for details.

Box 5. CUAHSI Community Modeling System

Models are used in hydrological science to integrate data with concepts and understanding so that conclusions about complex hydrologic processes can be drawn and quantified. There are a plethora of hydrologic models. However, many of these models have not been adequately tested, and the limits of their applicability have not been defined. Nevertheless, new models, model components, and solution techniques are being developed and published in the literature continually. Therefore CUAHSI can facilitate the advancement of hydrologic science through the development and support of a community modeling system with the following desired characteristics:

- The system must be able to treat each process and element of the hydrologic continuum as well as all of its interfaces with related sciences.
- The system must be modular such that various conceptualizations of processes and subprocesses important in hydrologic science can be easily and seamlessly interconnected to form models of complex hydrologic systems.
- The system must contain simple interfaces with the CUAHSI data retrieval system.
- The system must be supported by well-vetted, standardized data sets describing each of the important hydrologic and critically related processes at a variety of spatial and temporal scales so that the outputs of various model components can be subjected to robust testing.
- The system must be interfaced with effective data visualization software so that input data can be both screened and interpreted.
- The system must include consistency checks and a subsystem for error flagging and reporting for basic tenets such as the conservation of mass.
- The system must have the option of providing sensitivity analysis with respect to input-data accuracy, initial and boundary conditions, modular constructs, and solution techniques.
- The system must have the capability of generating error bands on output data values for static runs of the resulting model and should also be able to be used in a probabilistic forecasting mode.

The groundwater modeling community has settled to a large extent on a community model, MODFLOW, which could be incorporated as an initial element of the CUAHSI community modeling system. Furthermore, The U.S. Geological Survey has developed a robust Modular Modeling System for precipitation-runoff computation that could also provide an opportunity for synergy between the efforts of CUAHSI and the federal government.

HYDROLOGIC INFORMATION SCIENCE CENTER

The Need

The CUAHSI strategy for hydrologic research support hinges on three infrastructure initiatives: *Hydrologic Information Systems*, *Long Term Hydrologic Observatories,* and *Measurement Technology.* These three interacting elements make sense as part of a *Hydrologic Observing System* where observations on varied elements of the hydrologic system and the variables that influence it can be brought together, and the relative utility of existing and planned systems for measurement, modeling, and analysis can be assessed as part of a community research strategy (see Box 6). A natural place to achieve this synergy and coordination across the infrastructure elements is the Hydrologic Information Science Center (see Box 7 for examples of how this center might be used.

Hydrologic information science is built on scientific advancements in the field of informatics that facilitate the development of information technology necessary for advancements unique to hydrologic studies. Examples include improved hydrologic data representation for synthesis and analysis of information at a variety of spatial and temporal scales, paradigms for knowledge discovery to explore complex dependencies among various process elements, and development of metadata standards for the representation and distribution of hydrologic information. This requires a synergy among various disciplines such as hydrology, computer science, geographic information science, mathematics, atmospheric science, statistics, and engineering.

Such a multidisciplinary approach also warrants a place where researchers can work in groups to advance major problems and science questions, identify critical gaps in knowledge, and develop personal and community knowledge of the latest data, their interpretation, and associated knowledge representation frameworks. It is anticipated that this facility would be located at a research university, with satellite (possibly virtual) sites at other research universities, and would have a small administrative and support staff to facilitate conferences, workshops, technical training, education, and outreach, in addition to facilitating integrative research of students and visitors aimed at synthesizing available data and information.

Relation to the Science Plan

The realization of the CUAHSI Science Plan requires more than just data and tools; it depends most of all on a coherent intellectual foundation for the advancement of the science. This foundation must take account of the structures within which data resides and tools are created, but at a higher level it must add a vocabulary and language for describing new systems and structures, for understanding the relations among the concepts, and for formulating concepts mathematically in a coherent form before they are built into coding systems. For example, it is easy to talk in general terms about "spatial and temporal scales" of information, but information system design requires being absolutely precise about what these terms mean and this introduces many subtleties of definition and specification that are not obvious at the outset. Information interchange requires the design of metadata standards for hydrology, which requires a significant amount of debate within the community. Being able to address new comprehensive scientific questions within multi-institution project teams working together over periods of several years, requires infrastructure for these teams to meet, collaborate, and be supported in their activities.

Hydrologic information science is a branch of hydrologic science that advances knowledge through the synthesis of hydrologic information. Some branches of hydrology, such as sediment transport, have depended very largely on information synthesis throughout their history. Others, such as river flow routing, have depended more on physical laws and mathematical equations. New fields of inquiry such as transport of toxic substances from the Mississippi Basin into the Gulf of Mexico, depend on information synthesis over large geographic regions.

Hydrologic information science will be built on hydrologic information systems in much the same way that the field of geographic information science has been built on geographic information systems, and indeed has supplied some of the underpinnings for the advancement of those systems. The geographic community has formed a University Consortium for Geographic Information Science (UCGIS) (http://www.ucgis.org), which is a non-profit organization of universities and other research institutions dedicated to advancing the understanding of geographic processes and spatial relationships through improved theory, methods, technology, and data. The list of academic member institutions in UCGIS is similar to that of CUAHSI.

The advancement of hydrologic information science requires an environment for debate, enquiry, collaboration, and study.

How the Need will be Met

This need will be met by the development of a *Hydrologic Information Science Center* (HISC), which serves as a meeting place and collaborative research environment for hydrologic scientists to develop new information system prototypes, to define new data standards and methods for data set construction, and to interact with scientists in other fields such as computer science and atmospheric science. This component is also envisaged as being a physical facility located at or near a major university research institution. It may be located with the Hydrologic Data Access Center or located separately. There may also be merit to locating the center with other related national centers, such as the National Center for Ecological Analysis and Synthesis (NCEAS), the National Centers for Atmospheric Research (NCAR), or the National Center for Supercomputing Applications (NCSA). Because integrative science is at the heart of CUAHSI and not just its Hydrologic Information Systems component, this may become part of a similarly conceived center of larger scope supporting CUAHSI as a whole.

The Hydrologic Information Science Center would not itself be a funding center for financial support of the salaries of investigators on collaborative projects but rather would provide funding for meetings, and a centralized support infrastructure to facilitate teams of investigators meeting periodically over a period of perhaps two to three years to advance particular project goals.

Proposed Activities

The Hydrologic Information Science Center would include the following program elements:

- Hydrologic Information Technology Prototype Research. These are initiatives with the goal of developing prototypes for systems not yet in existence so that their feasibility for conversion to fully engineered CUAHSI HIS components can be properly examined.
- Definition of Hydrologic Metadata Standards. This will develop self-describing data by defining a vocabulary and grammar for hydrologic data in a hydrologic variant of the Extended Markup Language (XML) now becoming a standard for self-describing data in the software industry. This activity may usefully be coordinated with other related activities sponsored by the federal government, such as the Federal Geographic Data Committee.
- Hydrologic Information Science Initiatives. These would be activities undertaken with other science institutions such as the NCSA Globus project and the grid projects for distributed scientific collaboration.
- Hydrologic Benchmark Data Sets. This activity would encompass a small number of large-scale collaborative projects, such as the definition and construction of benchmark data sets for precipitation, evaporation, streamflow, and other hydrologic variables, and perhaps for their integration to construct a water balance of the United States or a global water balance.
- Information Science Support Services. This would provide technical support for research meetings organized at the center to evaluate existing modeling and observing systems and propose modifications or new programs. Access to CUAHSI and other information sources and support for the rapid deployment of statistical tools for evaluations would be provided. Because many such meetings would entail collaboration with other disciplines, the center would facilitate the integration with relevant information sources.

Box 6. Hydrologic Observing System

Hydrology encompasses many scales and diverse media, and interfaces with many disciplines. This diversity necessitates a systematic, integrative, and adaptive approach to hydrologic information organization and data collection campaigns. Traditionally, hydrology has been organized largely as a terrestrial science, with "sub-divisions" of surface and ground water. However, it is now clear that the dynamics of the entire global hydrologic cycle, connecting the atmosphere, ocean, cryosphere, biosphere, and land needs to be understood. Hydrologic fluxes exhibit organization and predictability across large spatial and temporal scales, and yet, we have limited predictability at the finest scales of analysis if local processes only are modeled while keeping the larger-scale fluxes as an exogenous forcing. CUAHSI's programs for research into the deployment of new sensors, and development of regional observatories need to be hierarchically organized as part of a plan to support discovery of process and pattern at multiple scales and across multiple variables, in addition to efforts directed at an improved understanding of hillslope and watershed scale processes.

A Hydrologic Observing System (HOS) is proposed as a synthesis function of the Hydrologic Information Science Center. At a basic level, the HOS facilitates real-time multi-resolution access to a diverse set of data. Thus, an investigator at an observatory or a university could quickly assess the current conditions at locally instrumented sites and their recent evolution relative to long term fluctuations in the region, and develop an understanding of how the local and regional hydrologic fluxes relate to those in other regions, and to the phase of ENSO. Such an intuitive understanding could facilitate the development of a new generation of models to explain the critical active processes in a certain context, instead of the operational linking of disciplinary modules developed and calibrated to data at certain resolutions. The identification of variables that need to be better sampled to improve our understanding of emergent phenomena across multiple scales would be a byproduct of such activity.

An example of an integrative activity that could be pursued under the HOS umbrella is the reconstruction of streamflow data for the past century or the past four centuries. Hydrology has a rich history of empirical studies to regionalize floods, impute monthly and annual flows, simulate daily weather, and generate runoff As we embark on high-resolution measurements, recognize the spatial and temporal teleconnections of global water fluxes, and become aware of paleo-record extension techniques, the streamflow reconstruction problem translates into an unprecedented opportunity to integrate multidisciplinary science with multi-resolution measurements (over land, air, water) to provide a mechanistic basis for reconstructing various attributes of streamflow. The reconstruction process would also allow us to separate the local effect of human influence from natural variability over the last century because global and local data fields would be used to constrain the analysis. Thus, the Hydrologic Observing System can help us look at what is happening today, understand the past, predict the future, and identify what new data to collect and where, while maintaining scientific and social relevance.

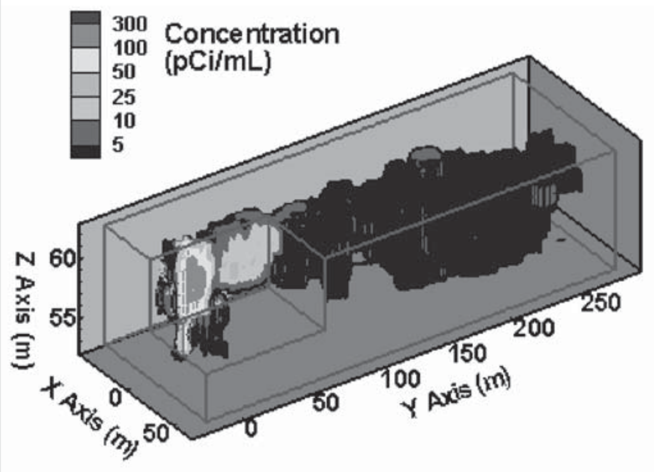
Box 7. Groundwater Data Sets for Model Testing

Field studies at well-instrumented sites have played a prominent role in efforts to better characterize and understand solute transport processes in groundwater systems. In particular, several recent field studies conducted at sites such as Borden, Cape Cod, and the Columbus Air Force Base in Mississippi, have provided new insights and comprehensive data sets for development and testing of solute transport theories and mathematical models.

As an example, consider the tracer experiment site located inside the Columbus Air Force Base (see figure, left). The Columbus site has been commonly referred to as the MADE (macrodispersion experiment) site and, more recently, as the NATS (natural attenuation study) site. Since the early 1980s, data collected at the Columbus site have been used by numerous researchers from the United States and around the world to gain insight into contaminant transport processes in highly heterogeneous aquifers and to develop new and improved theories and computer models to enable more accurate description and prediction of contaminant transport and remediation. Much of the current research on contaminant transport has been motivated by Columbus site findings that indicate small-scale heterogeneities and resulting preferential flow paths may exert dominant controls on contaminant transport. For instance, at least four hypotheses and alternative models have been proposed to explain the strongly non-Gaussian and other anomalous behaviors of the solute plume from an instantaneous source at the Columbus site (see figure, right). These include fractional advection-dispersion, non-local dispersion, correlated random walk, and dual-domain mass transfer. Testing and comparing these alternative models against existing and new field data sets is of fundamental importance and great interest to the hydrologic science community.

The Hydrologic Data Access Center and Hydrologic Information Science Center provide an ideal mechanism for collecting and archiving both comprehensive field data sets, such as those from the Borden and MADE sites, and smaller ones from individual researchers and research groups, and for making the data sets readily accessible to the larger community. The physical facilities associated with these centers provide a focal point for researchers of mutual interest to compare and explore alternative models and hypotheses.

The MADE site in Columbus, Mississippi.



Observed tritium plume from the MADE site.

RESOURCE REQUIREMENTS

The Hydrologic Information Systems component of CUAHSI requires resources for the Hydrologic Data Access Center (HDAC), the Hydrologic Information Technology Program (HITP), and the Hydrologic Information Science Center (HISC). We think that the third component, the Hydrologic Information Science Center, is fully justified within the mission of the HIS component of CUAHSI, but there is a larger sense of integration across all of CUAHSI's activities that a synthesis center could perform. A separate proposal has been developed within the CUAHSI structure for a Synthesis Center for Hydrologic Science & Technology (SCHiST), and to avoid confusion between these two proposals for synthesis centers, we have included the costs and personnel requirements anticipated for the SCHiST as part of this proposal, rather than trying to estimate costs and personnel for a standalone Hydrologic Information Science Center.

We have proposed a two-phase approach to developing CUAHSI Hydrologic Information Systems over a five-year period. Phase I would last for the first two years and would focus on startup activities to develop a working system for each of the three CUAHSI HIS components. The knowledge gained during this startup phase would then inform a set of Phase II activities over three to five years, which would expand and consolidate the CUAHSI HIS mission.

The anticipated resource requirements are focused on four major components: hardware, personnel, facilities, and external funds for contracts. The first three items will have an initial cost component as well as a continuation component to operate and maintain the facilities for a number of years. The external funding item may include just an initial amount of what would become a larger figure later because much of the outlined functionalities of the programs and activities do not yet exist and would need to be developed throughout Phase I. The staffing and facility needs for the Hydrologic Information Science Center (or Synthesis Center for Hydrologic Science and Technology) follow the lines outlined at the CUAHSI website.

HARDWARE FOR HDAC

The necessary hardware for the Hydrologic Data Access Center is defined around the needs specified by the user community and the necessary technology to serve these needs. As such, there are two types of hardware needs: (1) the HDAC server side configuration, and (2) staff office and development machines for development and maintenance. For the server side we anticipate a configuration that in its simplest form (there may be other or multiple components added to the layout later) is presented in Figure 5.

The exact configuration of these servers is subject to a number of factors that will need to be carefully worked out and subsequently monitored. Some of the most important aspects are the "hit" rate of users, data stream requirements for the "push" technology, the level of performance computing that is offered through the community model, the data storage requirements (CUAHSI-generated as well as external data), and the level of data processing (e.g., visualization, added-value data products, data mining, knowledge discovery, and grid data processing) that is deemed necessary to provide as a service to the user community. Because of these uncertainties, it is necessary to employ a number of monitoring efforts to deduce what exactly is needed in the mid- and long term.

Therefore, during Phase I, a reduced data-server configuration should be implemented to support basic needs. This

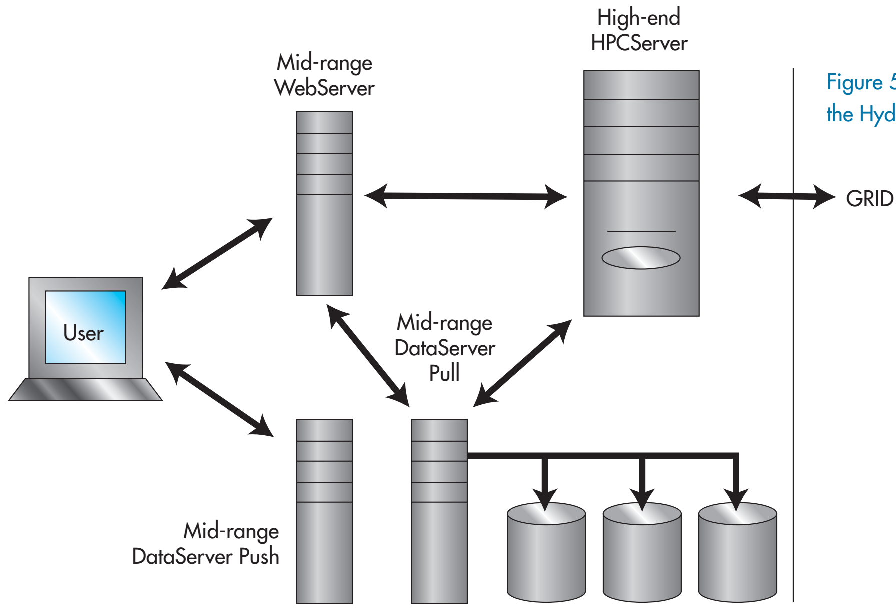


Figure 5. Hardware configuration for the Hydrologic Data Access Center.

should be sufficient for a quick launching of the HDAC website, providing low-level services like introduction of the various groups and initiatives, objectives, and perhaps basic download aids for data that already exists elsewhere. The site could be modeled on the current CUAHSI website. Phase I equipment for the server configuration should consist of three small servers that could later be reconfigured and integrated into the final configuration (Table 1). The office (computer lab) configuration is determined by the number of staff required for the operation of the HDAC/HISC (the projections are given below) and requires a combination of Unix and Windows machines for the staff. The variety of the machines is necessary to ensure that system administration and development can be tested on the most commonly used operating systems and its interoperability guaranteed.

Phase II will see an upgrade of the server side system for which it is anticipated that the initially installed servers will need to be upgraded to three mid-range size servers and one high-end size server (Table 1). The exact configuration will then be determined by the data gathered during the initial startup phase, which is anticipated to last for about two to three years. Also, a considerable amount of data storage has to be incorporated into the system. The exact amount is not crucial as storage can be added with the appropriate software licenses when need arises.

The costs in Table 1 are taken from current price lists of vendors such as Sun Microsystems, Silicon Graphics, IBM, and Dell, and do not vary much from vendor to vendor for a certain configuration. These numbers represent initial estimates and may not be complete. In addition, they most certainly will vary or change once the exact configurations emerge, but they should provide a ballpark estimate that can be used for future planning.

Table 1. Hardware needs and costs for CUAHSI Hydrologic Information Systems Phase I and Phase II.

Phase I	
3 small size UNIX servers @ $30,000	$90,000
Data Storage	$25,000
7 UNIX workstations @ $30,000	$210,000
9 WINDOWS machines @ $4,000	$36,000
Peripherals	$20,000
Miscellaneous	$10,000
	~ $400,000

Phase II	
3 mid-size UNIX servers @ $350,000	$1,050,000
1 high-end HPC server @ 1,000,000	$1,000,000
Data Storage	$700,000
7 UNIX workstations @ $30,000	$210,000
7 WINDOWS @ $4,000	$28,000
	~ $3,000,000

STAFFING PLAN FOR HDAC

The personnel requirements are reflected through the hardware and development needs identified earlier. Eight operating groups are necessary for the operation of the center sites:

- Computer System Admin (WINDOWS, UNIX) to keep machines running and maintain network.
- Data System Administration (metadata, data conversion formats, data distribution, "push" and "pull"), and activity surrounding data streaming.
- High-Performance Computing administration (HPC, grid, community model), to get the community model operable, and deal with HPC parallelism, grid, etc.
- WWW administration to keep the website functioning and develop applet servlet applications (the "JAVA" person).
- A tool box administrator for analysis tools, GIS applications, and visualization.
- A technical support person who provides telephone and Internet support services, answers the "hotline," a person that deals with "… how do I get …?"
- CUAHSI administration (coordinates with the other activities of CUAHSI, outreach needs, HISC coordination, etc.)
- Center Administration (director, finances, personnel, etc.)

Phase I would see one person each for these administrative units plus a center director. Phase II would see an upgrade (possibly a doubling of the head count) that may be comprised of additional specialists or graduate students. The increase in staffing will depend on the specific needs of each operating group and may include none in one and perhaps two in another.

Table 2. Staffing costs for CUAHSI Hydrologic Data Access Center Phase I and Phase II.

Phase I	
	$960,000/yr
Phase II	
$960,000 + 840,000	$1,560,000/yr

FACILITIES FOR HDAC

The number of staff and initial discussions indicate that a facility with ~ 3000 ft^2 needs to be established. Besides providing ample office space, it should also include a conference room(s), a computer room with individual climate control, and kitchen and bathroom facilities. Whether the space exists somewhere rent-free, or needs to be rented depends on the circumstances and will be addressed later. However, an initial rate of $7/ft^2/month will be assumed. In addition, irrespective of the rental situation, it is assumed that initial renovation costs, investments in infrastructure (high-speed data lines, telecommunications), as well as office furniture and office supplies will be necessary (Table 3). Finally, continued utility services, custodians, as well as general maintenance and operating costs need to be figured in on a yearly basis.

Table 3. Initial and ongoing facility operating costs for the CUAHSI Hydrologic Data Access Center.

Initial	
(renov., office supplies, infrastruc.)	$500,000
Operations	
(rent, utilities, maintenance)	$300,000/yr

EXTERNAL CONTRACT FUNDS HDAC AND HITP

Much of the technology that both needs to be employed for the operation of the HDAC and that is part of the research plan under the HITP must be developed during the early stages (i.e., Phase I, because it is not readily available). It is anticipated that a number of requests for proposals need to be formulated for contracts to address the various research and development needs. These may include:

1. Development of a metadata standard for hydrologic data that encompasses the heterogeneous data-source situation as well as the newly developed data product list, also referred to as "added-value-data." This task must have a very high priority as it forms the basis for almost all data operations to be performed by HDAC and HITP.
2. Development of a data toolbox to visualize, inspect, analyze, interpolate, and combine with other data sets in space and time. The extent of this toolbox must derive from the needs of the user community and may be provided as a WWW online tool or a set of applications that can be downloaded and combined with commercially available software.
3. Development of a community model(s) with benchmark data sets. This includes the framework by which the user community can contribute to the model as well as online simulation capabilities that also includes the use of the grid. This is a high-performance computing (HPC) application that will require considerable research and development efforts to implement.
4. Development of a functional and logical website through which all services are provided or can be subscribed to. This is a formidable task that reaches from graphic design of the site to functionality, operational logic, and security and must be coordinated with all other CUAHSI groups as well as the various development groups inside HDAC and HITP. This task is most important as it will be the access portal to the community; hence, it is at the frontline for gaining community acceptance.

Each of these tasks should be started at the beginning of the CUAHSI HDAC/HITP irrespective of where the centers will eventually be located. Each of these tasks will require a development time frame of three years, and each of these tasks may include a multi-PI research team. Also, there may be other research and development projects that need to be identified on a smaller scale. Considering the scope of each of these projects, an amount equivalent to $600,000-$700,000 must be allocated. In subsequent years (Phase II), more research and development needs may be identified as part of the HITP or through synergistic activities in the HISC for which funds should be made available as well. The scope of the continuing research, however, is unclear and can only be roughly estimated.

Table 4. External contract funds needed for research and development in Phase I and Phase II for the CUAHSI Hydrologic Data Access Center and Hydrologic Information Technology Program.

Phase I	
	$2,800,000
Phase II	
$960,000 + 840,000	$1,000,000/yr

BUDGET SUMMARY HDAC AND HITP

Table 5. Budget Summary for the CUAHSI Hydrologic Data Access Center and Hydrologic Information Technology Program. Phase I is years 1 and 2, Phase II is years 3-5.

Phase I		
	Fixed Costs	Annual Costs/Year
Hardware	$500,000	
Personnel		$1,000,000
Facilities	$500,000	$300,000
External Contracts		$1,400,000
Total	$1,000,000	$2,700,000

Phase II		
	Fixed Costs	Annual Costs/Year
Hardware	$3,000,000	
Personnel		$1,500,000
Facilities		$300,000
External Contracts		$1,000,000
Total	$3,000,000	$2,800,000

TIMELINE FOR HDAC AND HITP

After funds have been made available to CUAHSI, a solicitation process should start immediately. This includes foremost an announcement for hosting the center and should be accompanied by the announcements for the research projects outlined above. A tentative timeline is provided in Table 6.

Table 6. Timeline for solicitations for the CUAHSI Hydrologic Data Access Center and Hydrologic Information Technology Program.

Activity	Y1	Y2	Y3	Y4	Y5	Y6
Center Solicitation	***					
Research Projects Solicitation	***					
Center Facility Implementation	*****	*****				
Research Project 1	*****	********	********	***		
Research Project 2	*****	********	********	***		
Research Project 3	*****	********	********	***		
Research Project 4	*****	********	********	***		
Acquisition Hardware Phase I		**				
Recruiting Personnel Phase I		**				
Initial Site Implementation		***	*			
Implementation of Research Results				**	***	
Acquisition Hardware Phase II				**		
Recruiting Personnel Phase II				**		

HARDWARE FOR SCHiST

The necessary hardware for the Synthesis Center for Hydrologic Science and Technology (SCHiST) is defined around the usual computing needs of the center staff. No HPC is needed, but the equipment should enable the staff to adequately communicate with and through networks (Intranet and Internet). Table 7 lists the anticipated needs.

Table 7. Anticipated hardward needs and associated costs for the CUAHSI Synthesis Center for Hydrologic Science and Technology.

Phase I	
22 WINDOWS machines @ $5,000	$110,000
10 UNIX machines @ 20,000	$200,000
Peripherals	$30,000
Miscellaneous	$10,000
	~ $350,000

Phase II (upgrades every 4 years)	
22 WINDOWS machines @ $5,000	$110,000
10 UNIX machines @ 20,000	$200,000
Peripherals	$30,000
Miscellaneous	$10,000
	~ $350,000

STAFFING PLAN FOR SCHiST

Based on the experience of NCEAS, we estimate that the overall staffing requirements for SCHiST will range between 15 and 20 persons, who can be categorized as follows:

Management: The management group will be comprised of three persons: the director, the associate director, and a special assistant to the director. This group will be responsible for the overall management of the center, for planning its activities in conjunction with the SCHiST Advisory Council, for developing its sources of funding, and for networking with the hydrologic science community.

The most crucial member of the SCHiST team will be the director, who will set the tone of cooperation and mutual support that will be necessary if the concept is to succeed in the hydrologic science community. The director should be a known and respected member of the community that he or she will serve, and the director also must have the interpersonal skills and the scientific savvy to foster the interdisciplinary studies that will be the backbone of the center's scientific programs. On the other hand, the director must have the restraint such that he or she does not or is not perceived to control the agenda of the center to the detriment of the broader community. Ideally, the director also would have managerial experience so that both the day-to-day operation of the center would be effective, and its long-term vision would be guaranteed by the development of a high-quality business plan.

The associate director will serve as the primary assistant to the director, and will serve in the director's stead when the director is absent from the center. Because it is not likely that a director can be hired with all of the desired attributes of an outstanding caliber, an associate director who complements the director's strengths and weaknesses should be hired.

The special assistant will serve as the third-ranking official for the center, but he or she will have no supervisory responsibilities when either the director and/or the associate director are present. The primary role of the special assistant is to work closely with the scientists and the work groups to assure the success of their endeavors. This individual should have experience in group dynamics and also would serve as the primary liaison between management and the working groups and scholars.

Scientific Support: The Scientific Support Branch would be comprised of two specific groups, a Project Support Group (PSG) and a Hydroinformatics Group (HIG), which would work very closely together. The HIG would collaborate closely with activities of the Hydrologic Information Systems activities of CUAHSI.

The PSG would provide support for the projects of the center throughout the life of each project. It would facilitate the activities of the working groups, assist the resident scholars, prepare reports and outreach materials, and organize workshops, conferences, and training courses hosted by SCHiST. The PSG would be staffed with approximately six individuals with about half of the group having skills and training in research applications tools such as statistics, data visualization, and data base management. Another part of the group would be skilled in technical writing and would have responsibility for the preparation of workshop and conference reports and training and outreach materials. The leader of this group could come from either of the cohorts.

The HIG would have responsibility for identifying, retrieving, and collating the data and information required for each of the SCHiST projects. This group also would be responsible for the library facilities that may be established at SCHiST. A staff of approximately three persons would have backgrounds and capabilities in information science.

Technical Support: The Technical Support Branch will be responsible for operation, maintenance, repair, and upgrade of the center's software and hardware as well as operating the center's intranet. Secondarily, this branch would provide limited, but critical, computer programming for the Scientific Support Branch. The Technical Support Branch would require the services of approximately five individuals encompassing a broad array of experience in computer science and applications.

Support Services: Three to four secretarial positions would be required to support the activities contemplated for SCHiST. In addition, one clerical position would be needed to handle the travel arrangements for the participants in the multitude of programs envisioned. A bookkeeper would be needed to work with the accounting staff of CUAHSI and with the management of SCHiST to make sure that the center's finances were properly allocated and tallied. One custodian would be needed to keep the facilities functioning, clean, and orderly. Table 8 summarizes the anticipated SCHiST staffing requirements.

Table 8. Summary of anticipated SCHiST staffing requirements and costs for Phases I and II.

Management	1	Director
	1	Associate Director
	1	Special Assistant to the Director
Scientific Support Branch	Project Support Group	
	1	Group Leader
	3	Research Applications Specialists
	2	Technical Writers
	Information Group	
	2	Information Specialists
	1	Librarian
Technical Support Branch	1	Computer Scientist
	1	Network Administrator
	3	Computer Technicians
Support Services	3	Secretaries
	1	Travel Clerk
	1	Bookkeeper
	1	Custodian
Total	23	

Phase I	
	$2,000,000/yr

Phase II	
	$2,000,000/yr

FACILITIES FOR SCHiST

The physical facilities required to house the SCHiST would consist of a building, its furnishings, and equipment.

NCEAS, with an operation not very different than the one described above, is located in an office building in downtown Santa Barbara, California, occupying almost 11,000 square feet of the building. In contrast, a workshop that was convened to scope out the needs of NCEAS before it was established estimated that approximately 20,000 square feet would be needed to carry out the mission that it espoused. The main difference between the workshop estimate and the actual facility is the inclusion of a 100-seat auditorium in the plan that did not come to fruition in the real world. Thus, for initial planning purposes, a building of approximately 8,000 square feet will be assumed to satisfy the needs of SCHiST.

The facility itself should be designed or remodeled so that the natural interaction of resident scholars and visitors is encouraged while still offering solitude for those times when the creative process needs to occur without interruption.

The location of SCHiST will be a major determinant of its success. SCHiST should be located near a university campus with a major program in hydrologic science so that natural synergies between the university and the center can be fostered. Some of the possible synergies will involve the participation of students, both undergraduate and graduate, in the activities of the center either in work study programs or as junior participants in the SCHiST working groups. The university's faculty also will benefit from ease of access to the center programs and from collaboration with its resident scholars. The center will benefit from the interaction with the local university faculty and through sharing of facilities such as libraries and internet capacities. The ideal locale would be one where the local university currently is a node on Internet II or soon will be for access to this latest expansion in bandwidth could offer significant advantages to SCHiST.

On the other hand, the center should not be located on a university campus so that perceptions of SCHiST being a captive of the university's agenda for hydrologic science would be minimized. Isolation from the actual campus also provides some insulation from the day-to-day routine for the local faculty when they participate in center activities. The fact that NCEAS is located some miles from the University of California at Santa Barbara has been shown to be a definite attribute.

Other locale characteristics that would aid in attaining the center's mission would be that it is located in a town where the climate and the ambiance are pleasant to visit and to live. The climate should not be too extreme in any season as the enter will be in operation year round. The local scene should offer pleasant and reasonably priced housing for the resident scholars and the staff and have comfortable and cost effective lodging for the transient participants in the work groups and workshops as well as a reasonable choice in cuisine among the local restaurants.

Lastly, the center should not be so isolated from the air travel service that it deters participation in work groups and workshops by scholars located throughout the nation and the world.

The center should be furnished in a manner that is pleasant, functional, and comfortable, with adequate space and storage for those who will spend significant portions of their time therein. Each office and conference room, as well as the lounge, should have adequate equipment to serve as a stimulus to accomplishing the SCHiST mission. Once again, NCEAS can serve as an initial model for the SCHiST concept. NCEAS has adequate computers of various types such that each visiting scholar has access to his or her operating system of choice in his or her office space. Furthermore, NCEAS has two computer laboratories where resident scholars and visitors have access to a wide variety of specialized machines. Each of the NCEAS conference rooms is networked into the LAN with multiple sockets adequate for full participation of work group members.

Finally, the NCEAS lounge is a focal point of the interactions among those who participate in NCEAS activities. It is laid out so that several discussion groups can be active at any given time with comfortable furnishings and access to white boards for explaining and recording ideas.

Specific budget estimates for the start-up costs and the annual operating budget should be developed as more details of the concept are defined. For the purposes of this exposition, the cost estimates that were derived in the workshop to scope out the NCEAS facility and their comparison to the actual costs to date of the NCEAS operation should serve as a reasonable metric for expected costs of SCHiST. The NCEAS planning workshop estimated that the physical start-up costs for the initiation of NCEAS would be between $4.8 and $6.9 million, and that the annual operating costs would be between $3.7 and $7.0 million. These figures assume that the NCEAS building would be built and capitalized in the initial year. However, by leasing instead of building, the start up costs could be reduced to about $1.75 million, but the annual operating costs would have to be increased by about $800,000 to account for the lease payment (Table 9). In actuality, NCEAS does lease its facilities in Santa Barbara, and its start up costs in 1994 were $500,000; its current annual operating budget is about $2.65 million. We assume that the necessary facility will be leased.

Table 9. Summary of anticipated start-up and operating costs for the SCHiST facility for Phases I and II.

Initial	
(renov., office supplies, infrastruc.)	$500,000
Operations	
(rent, utilities, maintenance)	$1,000,000/yr

BUDGET SUMMARY FOR SCHiST

Table 10. Budget summary for SCHiST. Phase I is years 1 and 2; Phase II is years 3-5.

Phase I		
	Fixed Costs	Annual Costs
Hardware	$350,000	
Personnel		$2,000,000
Facilities	$500,000	$1,000,000
Total	$850,000	$3,000,000
Phase II		
	Fixed Costs	Annual Costs
Hardware	$350,000	
Personnel		$2,000,000
Facilities		$1,000,000
Total	$350,000	$3,000,000

Table 11. Total Budget Breakdown per Year for CUAHSI Hydrologic Information Systems (HDAC + HTIP + SCHiST)

	Annual Cost ($M)	**Allocation HDAC**		**Allocation HITP**		**Allocation SCHiST**	
		%	**$M**	**%**	**$M**	**%**	**$M**
Year 1	7.55	30.5%	2.3	18.5	1.4	51.0	3.8
Year 2	5.70	23.0	1.3	24.5	1.4	52.5	3.0
Year 3	9.15	52.5	4.8	11.0	1.0	36.5	3.4
Year 4	5.80	31.0	1.8	17.0	1.0	52.0	3.0
Year 5	5.80	31.0	1.8	17.0	1.0	52.0	3.0
Total	34.00		12.0		5.8		16.2

SUMMARY

The CUAHSI Hydrologic Information System proposal envisions the expenditure of $34 million over a five-year period. This will result in the creation of an entirely new infrastructure for developing hydrologic information science as a field of intellectual inquiry. There is currently no system of this type, either in the academic realm of hydrologic science or in the realm of hydrologic practice. This infrastructure will meet the widely recognized need of hydrologic scientists for better access to well organized hydrologic data. The CUAHSI Science Plan places a great emphasis on integration between hydrologic process, among scales of space and time, between hydrology and ecology, and between hydrologic science and water resources management. This integration requires an information infrastructure that does not exist and must be developed by CUAHSI. In part, the development of this HIS infrastructure is an intellectual task because building a comprehensive information picture must be based on logical principles from hydrologic science. In part, the HIS infrastructure development is a task of computers and software development, in order to provide the tools for data access, enquiry, assimilation and analysis that hydrologic scientists need. HIS infrastructure development also calls for new benchmark data sets and for development of metadata standards for describing hydrologic data sets.

The hallmark of any successful scientific organization is access to information relevant for proving or disproving hypotheses. The HIS component of CUAHSI will support:

- A Hydrologic Data Access Center for supporting hydrologic science through better access to hydrologic data.
- A Hydrologic Information Technology Program, which will create tools for hydrologic data analysis, assimilation, modeling, and information synthesis.
- A Hydrologic Information Science Center for fostering intellectual enquiry using hydrologic information and for definition of metadata standards, research information system prototypes, and benchmark data sets.

The infrastructure created by the CUAHSI Hydrologic Information Systems will support the research of hydrologic scientists throughout the nation. It will lead to the development of a new body of knowledge in hydrologic information science. It will be an integrating mechanism and guide for the application of hydrologic information and modeling in water resources management in the United States. In this sense, CUAHSI Hydrologic Information Systems is not simply a resource for hydrologic science, but an asset for improving understanding and management of the water resources of the nation.

This material is based upon work supported by the National Science Foundation under Grant No. 02-33842. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

CUAHSI

universities allied for water research

CONSORTIUM OF UNIVERSITIES FOR THE ADVANCEMENT OF HYDROLOGIC SCIENCE, INC.

TECHNICAL REPORT #2

AUGUST 2002