

Project Description

PROJECT DESCRIPTION

Mission statement

We propose to develop a cyberinfrastructure framework to support large-scale water quality data integration, analyses, and visualization in the Upper Mississippi River Basin (UMRB) *in real time* using data-enabled information technologies. Seamless integration of all existing real-time and ad-hoc monitoring efforts, and continuous modeling and data resources is a major challenge in big data context. This project is only possible within the framework of a BD Hubs and Spokes ecosystem because it requires significant: 1) knowledge and collection of water quality data from a wide range of academic, agency, and NGO sources across several states; 2) integration of data (of varying quality, format, and duration) into a single user-friendly system; 3) input from partners and stakeholders to understand the great variety of ways in which the data may be best accessed and utilized; and 4) computing resources. Partners include researchers at the University of Iowa (IIHR-Hydroscience & Engineering), the University of Illinois Urbana-Champaign (Great Lakes to Gulf Virtual Observatory - GLTG, National Center for Supercomputing Applications - NCSA), Iowa State University, and the Lewis and Clark Community College (National Great Rivers Research and Education Center). The team will collaborate with the Midwest Big Data Hub's Food-Water-Energy and Digital Agriculture spokes.

Introduction and background

Mobilization and delivery of nutrients (nitrogen and phosphorous) from point sources and farmed fields to the UMRB stream network is a decades-long problem (Baker et al., 1975; Burkart and James, 1999; Turner et al., 2008; David et al., 2010). In particular, seasonal Gulf of Mexico hypoxia caused by nutrient pollution delivered via the Mississippi and Atchafalaya Rivers and their tributaries is a problem that seemingly defies solution (National Research Council, 2008). Consequences include eutrophication of local and regional water resources (Turner and Rabalais, 1994; Mueller and Helsel, 1996; Dodds and Welch, 2000) and drinking water impairment (Weyer et al., 2001; Jones et al., 2016). Hydrological modification and pollutant loading in the region resulting from agriculture and other sources has had national consequences. Reduction of the anoxic area (hypoxia) in the Gulf of Mexico has been a national priority for over 20 years.

The Mississippi River/Gulf of Mexico Watershed Nutrient Task Force was formed in 1997 to coordinate an effort to understand and mitigate Gulf hypoxia (USEPA, 2017). An Action Plan was released by the Task Force in 2001 to serve as a strategy for hypoxic area reduction in the Gulf. A revised plan was released in 2008, which 12 states within the Mississippi Basin continue to implement. The Task Force's long-term goal then was to reduce the hypoxic area to 5,000 km² by 2015. Because the 5-year average size of the hypoxic area has remained largely unchanged since 1994, the goal was extended to 2035. Stemming the loss of nitrate-nitrogen from row crop areas has been an especially difficult problem. The 2001 Action Plan estimated that nitrogen loads would need 30 percent reduction to reach the hypoxic area objective; later research showed nitrogen reductions as high as 45 percent may be needed (Justic et al., 2003; Scavia et al., 2003; Rabotyagov et al., 2014a). Because NO_x-N delivery to streams is from a myriad of widely dispersed sources, including farm field drainage pipes and shallow groundwater (Baker et al., 1975; Burkart and James, 1999), regulations governing its release to the environment are nearly non-existent. As such, NO_x-N load reductions have been dependent upon education of, monetary incentives to, and voluntary actions of farmers in the region (Rabotyagov et al., 2014a). This approach has not manifested in reduced NO_x-N loading to the Mississippi River stream network (Sprague et al., 2011). In fact, the 2017 hypoxic area is reported to be the largest ever.

In response to this lack of progress, several states in the UMRB have instituted nutrient loss reduction programs of their own (Anderson et al., 2016; Illinois Nutrient Loss Reduction Strategy, 2014; Iowa State University, 2013). By adhering to strategies with stated objectives (i.e. 45% reduction), states have inherently made accountability part of the process required for the expenditure of public funds. Quantifying and tracking changes in nutrient delivery to the basin's stream network are key; many stakeholders are insisting that progress toward achieving reductions be quantified on a regular basis (Schilling et al., 2017).



Figure 1: Vision and Structure of the project

Strategic and scientifically-credible monitoring is the best way to track progress toward water quality objectives. The quantity of nitrate leaving Iowa is particularly well documented as Iowa has a statewide network of ~75 real-time, continuous nitrate sensors co-located with river discharge measurements. Data from all of these sensors is transmitted to the Iowa Water Quality Information System (IWQIS), which is the established mechanism for tracking nitrate loads in Iowa. The IWQIS visualization platform provides the public immediate access to this credible water quality data. Expansion of this network to the entire UMRB would provide multiple benefits to producers and land managers, scientists, policy-makers, municipal governments, agencies, and others seeking solutions to these difficult water quality challenges. By defining and implementing data and semantics specifications, and web service API (Application Program Interface), the expansion will be guaranteed to be interoperable with other data systems.

Approach

Our capacity to generate water quality and quantity data is exponentially exceeding stakeholders' capacity to digest and utilize it to manifest improvements from the field to the landscape scale. Key reasons for this include: 1) an unawareness of large but disparate data sources; and 2) a lack of accessibility to the data. *We hypothesize that creation of an easily accessible platform hosting regional and local water quality and quantity data, conservation practice effectiveness, climate and weather data and projections, and watershed information and other similar tools, will facilitate both high-level scientific discovery and easy access and utilization of data at the scales necessary to affect change that improves the condition of the region's water resources.*

The proposed Upper Mississippi Information System (UMIS) platform will provide novel data analytics and visualization capabilities to work on solutions for and monitor progress toward achieving the goals of the Gulf of Mexico Hypoxia Task Force. The platform addresses big data needs of states' (e.g., Iowa, Illinois, Wisconsin, Minnesota) nutrient reduction strategies through access to real-time (IHR, U.S. Geological Survey, U.S. Department of Agriculture, U.S. Army Corps of Engineers, and others) and historical (EPA STORET) data streams, which are all currently managed independently. The platform will enable data-driven visual analysis techniques to explore data derived through observation, simulation, citizen-science data collection, and user interaction to improve our understanding of water quality and quantity dynamics, as well as our ability to predict and warn communities of impending threats. The frequency and severity of these threats will only increase concurrently with the increasing incidence and magnitude of extreme weather events associated with climate change.

Potential research questions that could originate based on the proposed UMIS include:

- How could projected climate changes affect water quality from the UMRB to the Gulf of Mexico?
- What conservation practices are more effective in reducing water quality issues?
- Are the goals of the Gulf of Mexico Hypoxia Task Force achievable? If so, what will it take to achieve them?
- What is the most effective way to communicate water quality and quantity information to facilitate decision-making processes?

Study Region



The UMRB drains 189,000 square miles with most of this area lying within the states of Minnesota, Wisconsin, Iowa, Illinois, and Missouri (Figure 2). There are 279,000 farms operating in the basin where corn and soybean production occupy 44% of the watershed. Livestock populations include nine million cattle, 25 million hogs, and 100 million chickens. The area is a large contributor to the world supply of corn, soybeans, and animal protein, and the region has helped make the U.S. the world's largest producer of corn since the United Nations began keeping these records in 1961. Agricultural exports are an important part of the region's economy. This intense production has had environmental consequences, specifically soil erosion, eutrophication of local waters, and Gulf of Mexico hypoxia. Tributaries to the Mississippi River transport large amounts of nutrients, sediment, and pesticides, reducing biodiversity, impairing municipal drinking water supplies, and degrading the aesthetic and recreational value of water resources. Much of the natural hydrology has also been intentionally altered through the use of constructed drainage systems that lower the water table to make soil conditions more favorable for crop production. Very little of the native tallgrass prairie and wetland ecosystems remain. Replacement of these perennial systems with annual crops has also affected hydrology and water balances. The Mississippi River itself has also been radically altered with the lock and dam system to facilitate navigation so crops and other goods can be transported from the basin to other states and countries.

Proposed Research Tasks

The research tasks in the project are related to integration of existing real-time sensor network and ad-hoc citizen datasets, modeling activities and model repository, communication and visualization of models and data on existing cyber systems, and community building and training activities.

Task T1: *Integration of existing real-time water quality monitoring networks and crowdsource citizen science datasets by designing a data specification and communication framework.* Team members: Jones (lead), Demir, Lee, Warner, Weber, Krajewski)

Iowans have made accountability part of the process for the expenditure of public funds by proposing a strategy with a stated objective (45 percent reduction). Monitoring, quantifying and tracking the changes in nitrate delivery to the state's water resources is key for state and federal agencies, agricultural interests, environmental organizations, and many other stakeholders. However, quantifying nitrate load reductions generated by conservation practices is confounded by the region's dynamic weather patterns. Previous research has shown that NOx-N flux in the Mississippi Basin is strongly linked to streamflow and precipitation. Predicting NOx-N loss from agricultural landscapes is often not straightforward and can vary greatly between storm events, baseflows, and within small geographical areas. This is because the supply of NOx-N on the landscape varies seasonally, and because the dynamics of stream NOx-N flux are dependent upon antecedent moisture, season of the year and changing hydrologic pathways. For these reasons, water quality monitoring that includes nitrate as a measured parameter is critically important.

The Iowa Water Quality Information System (IWQIS) was created to generate credible water-quality data that would enable scientists, policy-makers, and the general public to track conditions in Iowa streams and subsequently inform implementation programs designed to improve stream conditions. A network of about 75 real-time water quality sensors continuously measures nitrate, turbidity, pH, dissolved oxygen, specific conductance, temperature, and discharge across the state. The sensor network and the companion web tool thus are an effective indicator of local and regional water quality, weather, and land-use practices. This system, however, is confined to water bodies within the state of Iowa and thus is not a predictor or indicator of Gulf of Mexico hypoxia.

Other data sets exist in Iowa and other states that could help characterize Mississippi River Basin water quality. These data are generated by traditional water monitoring programs conducted by agencies, environmental groups, agricultural advocacy organizations, and others. Furthermore, as technology for automated monitoring systems continues to mature and become cost competitive, implementation and integration of these devices into monitoring protocols and strategies will become more frequent. The capacity to generate data already outstrips society's ability to process these data, which could affect and track water quality and land uses changes. Integration of these other existing and future data sets into an UMRB-wide data tool would be invaluable to scientists, watershed coordinators, land managers and farmers, and policy makers as they work toward improving the condition of the region's water resources.

Challenges of integration of various water quality data include: 1) semantic differences with same parameters and measurements; 2) various data format /structure (e.g. time stamp format); and 3) various access mechanism. In order to overcome these challenges, specifications of data/semantics and web service API will be developed.

Research tasks related to data integration component include:

- Harvest of water quality data from the EPA STORET data base. EPA created the Storage and Retrieval (STORET) database to provide a public repository of water quality data. We will target data from states within UMRB (i.e. MN, WI, IL, and IA) that are critical for land management and water quality improvement measures in predominantly agricultural watersheds. This data will be harvested and featured on the UMIS in ways that help characterize and track water quality on the watershed and landscape scale within the agricultural UMRB.
- Identification and collection of state agency water quality and conservation practice data. This includes data from the departments of agriculture and natural resources in the four states. This data will be featured on the UMIS and will highlight water quality improvements that result from practice implementation.
- Identification and collection of water quality and conservation data from NGOs. Several farm advocacy groups and NGOs in the basin (Iowa Soybean Association, Illinois Council on Best Management Practices, The Nature Conservancy) are working with farmers to implement conservation and measure how these efforts are manifested in water quality data. We will work with these groups to highlight conservation success stories on UMIS, both with actual data and narrative.
- Collaborate with academic programs to generate and post water quality and conservation data. University extensions in some states are working with farmers to assess best management practice effects on water quality. For example, Discovery Farms (University of Wisconsin Extension) is working with farmers in both Minnesota and Wisconsin to measure water volumes and flows associated with row crop and animal agriculture. We also intend to collaborate with individual researchers working at the nexus of water quality and farming in ways that will inform practice implementation across the basin using water quality data generated at the field and practice scale.
- Work with individual farmers to measure water quality. We will work with farmer contacts to deploy equipment and water testing kits to measure water quality on the farm, with the intention that this data will inform conservation efforts via posting on the UMIS website.
- Collection and posting of data from the Iowa Watershed Approach project. The Iowa Watershed Approach (IWA) is a vision for Iowa's future that voluntarily engages stakeholders throughout the watershed to achieve common goals, while moving toward a more resilient state. This is a \$97M project funded by the U.S. Housing and Urban Development (HUD) in nine Iowa watersheds. Water quality and quantity data will be collected in each watershed that assess the success of practice implementation. This data will be posted on UMIS.
- Developing specifications for data and semantics on nutrient data. This task includes adopting and harmonizing existing data specifications (working with CUAHSI, EPA, USGS, and other agencies), and working with collaborators to define vocabularies on parameters (e.g. Nitrogen, etc.), collection methods, QA/QC, etc. Data format specification will support many output formats including but not limited to JSON, XML, CSV. Metadata specification will be used for sharing and cataloging purposes.
- Developing API specification for data access and query catalog. A web service API specification for community and interoperability will be developed by using data specifications previous task. The web service API will be used for querying the catalog (metadata) and data access (download).

Task T2: *Setting up a large-scale integrated modeling framework for continuous modeling activities at the Upper Mississippi River Basin.* Team members: Gassman (lead), Schilling, Demir, Weber, Krajewski

Our research team has developed a large, spatially-detailed, integrated modeling system for the UMRB and Ohio-Tennessee River Basin (OTRB) regions (Panagopoulos et al., 2015a; Kling et al., 2014) by configuring the Soil and Water Assessment Tool (SWAT) ecohydrological model (Arnold et al., 1998; 2005; 2012; Gassman et al., 2007; 2014) at the 12-digit watershed scale (USGS, 2013) in both basins. The current SWAT modeling system developed for the UMRB and OTRB regions is characterized by a higher resolution of subwatersheds relative to similar modeling systems developed for the two regions (Panagopoulos et al., 2015a) and has been previously used for analyzing the impacts of future climate change projections, pollution mitigation strategies or biofuel cropping systems (Panagopoulos et al., 2014; 2015a; 2017; Kling et al., 2014). The proposed activity here will focus on application of the UMRB SWAT model, which drains most of the western Corn Belt region and is a primary source region of nutrients exported to the Gulf of Mexico (Porter et al., 2015; USEPA, 2008).

Further spatial refinements will be incorporated by delineating more detailed hydrologic response units (HRUs) within each SWAT subwatershed to take advantage of available higher resolution soil, topographic, and land use data. These refinements will allow incorporation of the "landscape simulation concept" a new option that will be available in future SWAT codes (Bieger et al., 2017) that will allow improved spatial representation of conservation practices and land use. The HRUs will be based on the topographic, soil, and subsurface tile drainage criteria described in the classification scheme of Schilling et al. (2015). Refinements will also be introduced to more accurately represent subsurface tile drainage management, based on improved algorithms introduced in recent SWAT codes (Moriassi et al., 2013a, 2013b) that require further testing for the UMRB region. The climate data required to execute the SWAT models will be extended through 2017 to support baseline hydrologic and pollutant testing for the updated UMRB SWAT model. Software will also be developed to support real-time analysis of monthly SWAT outputs as a function of continuously updated climate inputs. Output from the monthly SWAT model runs for the UMRB will be databased as total nutrient loads exported from HUC12 watershed. Maps of the monthly nutrient loading patterns in the UMRB will be indexed and made available for display.

Research tasks related to data integration component include:

- Set up a large scale continuous modeling framework. The framework will be based on UMRB SWAT model and provide monthly model runs for the UMRB.
- Define a model specification and model repository. Develop a model specification and repository (working with CUAHSI, Hydroshare teams) that will be used for creating a model catalog and discovery service with metadata.
- Populate model repository. This will include modeling activities in the UMRB by any researchers, partners, and collaborators. The repository will support many model outputs (e.g. report, article, simulation results) and specifications to enable access to any stakeholders for evaluating the model results, use in planning and decision making, and reproducibility of science.

Task T3: *Develop new integration and communication tools for datasets, models, and services.* Team members: Demir (lead), Lee, Krajewski

The proposed project will benefit from two existing and operational information systems, Iowa Water Quality Information System (IWQIS) and Great Lakes to Gulf Virtual Observatory (GLTG). Scale and capabilities of IWQIS and GLTG will be extended to support the UMRB, and **Upper Mississippi Information System (UMIS)** will be developed based on this infrastructure. Real-time and ad-hoc datasets and model repository will be implemented to the UMIS and GLTG. An interactive platform will be developed to access and visualize model output for decision making with scenario-based model evaluation, data analytics and exploration, and geo-spatial visualization capabilities. The system will provide a desktop-like environment to manage, visualize and analyze rich high-volume geo-spatial data using web-based mapping and visualization capabilities. The implementation of web service API defined from Task T1 for UMIS and GLTG will demonstrate the interoperability among different information systems. Moreover, development of client tool for the web service will be helpful for community users to access the data with consistent manner.

The IWQIS is based on a generalized water cyberinfrastructure called Iowa Water Information System (IWIS). IIHR developed and maintains IWQIS as an easy-to-use web tool that harvests data from water-quality measurement devices (i.e., Hydrolab sonde) deployed in streams across Iowa. IWQIS is the platform for public dissemination of water-quality data from IIHR's network of water-quality sensors deployed in streams across Iowa. The IWQIS design is based on the vision and technical expertise of the generalized water cyberinfrastructure developed at the Iowa Flood Center (Demir and Krajewski, 2013; Krajewski et al., 2017). IWQIS provides an integrated and scalable workflow environment to support end-to-end cyberinfrastructure for sharing scientific research and observations (Gil et al., 2016; Duffy et al., 2012), as well as scientific visualization, communication of observations and modeling results that assist watershed management, and research (Demir et al., 2009; Demir and Beck, 2009; Demir et al., 2015).

IWQIS is optimized for the general public and is freely accessible to anyone with access to an internet browser. The platform is built on the easy-to-use and familiar Google Maps platform, and users can easily navigate and zoom to areas of interest. They can select any sensor and view current concentration or load, and view a graph of recent and annual trends. IWQIS is a one-stop platform (Figure 3) for water-related information in Iowa that provides data on precipitation, stream conditions, and many geospatial layers. IWQIS connects data from seven NEXRAD radars and more than 500 sensors in Iowa for rain, soil moisture, streamflow, and reservoir data. The information in IWQIS is organized around river communities; users can see watershed boundaries and river networks for more than 1,000 communities. With these features in place, IWQIS users can view water-quality data within the hydrologic context.

The GLTG is an interactive geospatial application that integrates water quality data from different sources to better facilitate analysis and provide knowledge of nutrient pollution and water quality conditions in the Mississippi River watershed. The National Great Rivers Research and Education Center (NGRREC) and National Center for Supercomputing Applications (NCSA) developed the Great Lakes to Gulf Observatory (GLTG) to connect environmental data with scientific concepts and understanding. The GLTG's web-based application (Figure 3) allows users to dynamically search for and visualize water quality information on the Mississippi River and its tributaries. The application uses a combination of different data sources including long-term data sets, providing the users with a better picture of water quality in the watershed.

The application's features, visualizations and design allows researchers, communities, and decision makers to better understand nutrient inputs and loads, expedite data-to-knowledge-to policy connections, enhance risk management, empower emergency response and inform long-term strategic planning. Data integration and visualization occurs within a cyber-infrastructure framework constructed in collaboration with the National Center for Supercomputing Applications (NCSA) and Illinois-Indiana Sea Grant (IISG) at the University of Illinois.

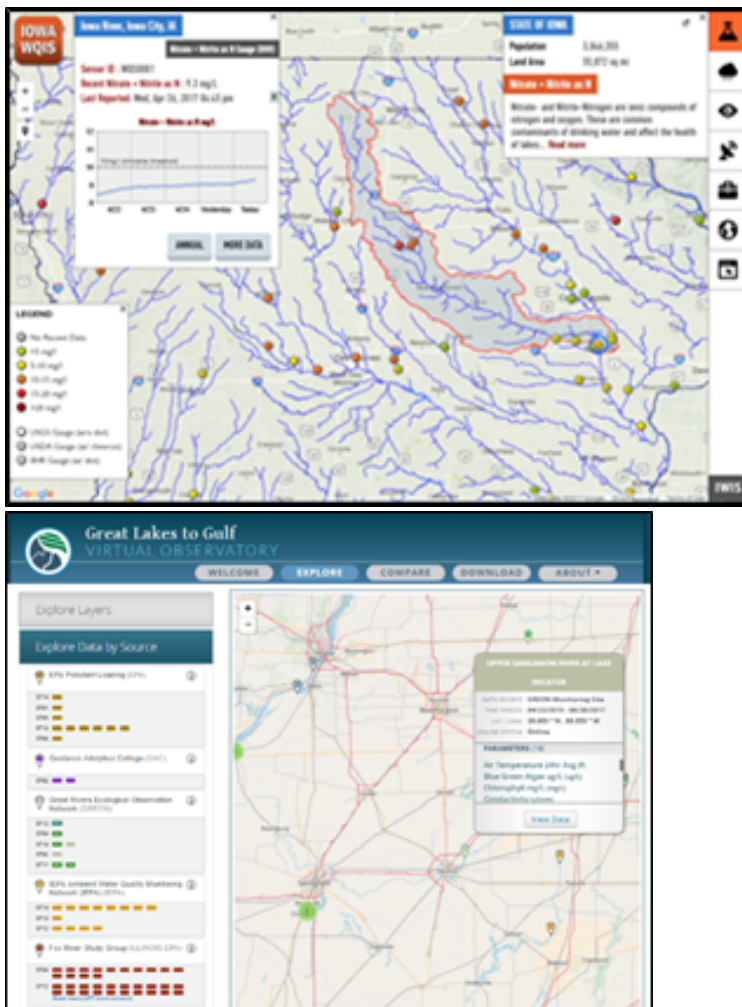


Figure 3. The IWQIS map interface with watershed and recent data graphs for a water-quality sensor (left), GLTG explore page (right)

Research tasks related to data integration component include:

- Extending scale and capabilities for IWQIS and GLTG to support new domain. Scale and capabilities of IWQIS and GLTG will be extended to support the Upper Mississippi River basin. A new information system, UMIS, will be designed to support Upper Mississippi River basin domain.

- **Web service implementation.** The developed web services (in Task T1) will be integrated to both UMIS and GLTG. Implementing web services on each system will demonstrate the interoperability in terms of end users and community. This allows end users to access data with same API regardless of system. At the same time, users can choose different User Interface if they want to browse the data among UMIS and GLTG. But they will have consistent data in terms of semantics, format and metadata.
- **Developing a client library or tools.** New client library tools (in python or R) will be developed to use web services from both UMIS and GLTG. This demonstrates the interoperability of systems, which follow data specification and API specification. The library (or tool) allows end users to query and access data from both systems at the same time.
- **Developing new visualization and communication tools.** For additional data and model products integrated to both systems, new visualization and communication tools will be developed. This includes tools for visualization of continuous and historical model runs, and real-time and ad-hoc data streams.

Task T4: *Collaboration with partners and education and outreach activities.* Full team.

This task includes three primary activities that will help to grow the UMIS and train new users.

- **Year one workshop:** Upon notice of project award, the team will begin preparations for a collaborator workshop in Iowa City, Iowa. About 50 people will represent as many relevant organizations as possible (see Roles and Responsibilities section for workshop commitments to date and likely invitations). The specific goals of this workshop are to seek input on the UMIS design and functionality, to survey participants regarding water-quality and hydrologic data sources, to solicit commitments from participants to share their data, and to open new long-term channels of communication between participants. The project budget includes participant support funds to cover workshop participants' travel and subsistence costs.
- **Partner-building activities:** After the workshop, the team will continue to identify new partners and resources (see Long-Term Sustainability section) to ensure the UMIS' continued growth and longevity, and to seek input on the tool's functionality. The team will especially leverage the MBDH resources to help grow the tool and find new partners and data sources. PI Krajewski and others will attend the MBDH "all hands" annual meeting to directly engage with researchers representing all of the hub's spokes. The team will prepare information to share with the MBDH for distribution through the hub's communication channels and for assistance in taking this project to the full basin or national level.
- **Outreach and training:** In project year three, researchers will prepare training activities for the UMIS. IIHR's successful training mechanisms for the IWQIS include webinars, videos (posted online to readily accessible sites like YouTube), and workshops/training events. All three of these mechanisms will be utilized. Costs will be minimal as the partners will use existing resources for the webinars and videos; workshops/training events will usually coincide with other events (watershed meetings, MWDH all-hands meetings, state and regional professional meetings, agency meetings, etc.), which will also maximize attendance.

Timeline

The project will span three years. In year one, the project will include close coordination and collaboration among partners, collaborators, and existing MBDH Spokes and projects for understanding the big data and cyber needs and requirements to address water quality challenges in the UMRB. This will include an extensive survey on existing data and technical resources and a year one workshop. We will start design and development of the cyberinfrastructure in the first year and start integrating existing data resources and capabilities. The development will continue until the end of the second year, and continue on refinements and improvements in the third year. We will have a pilot launch and collect feedback from partners in the second year. We will focus on training and support in the third year via workshops and webinars, and expand user base of the platform.

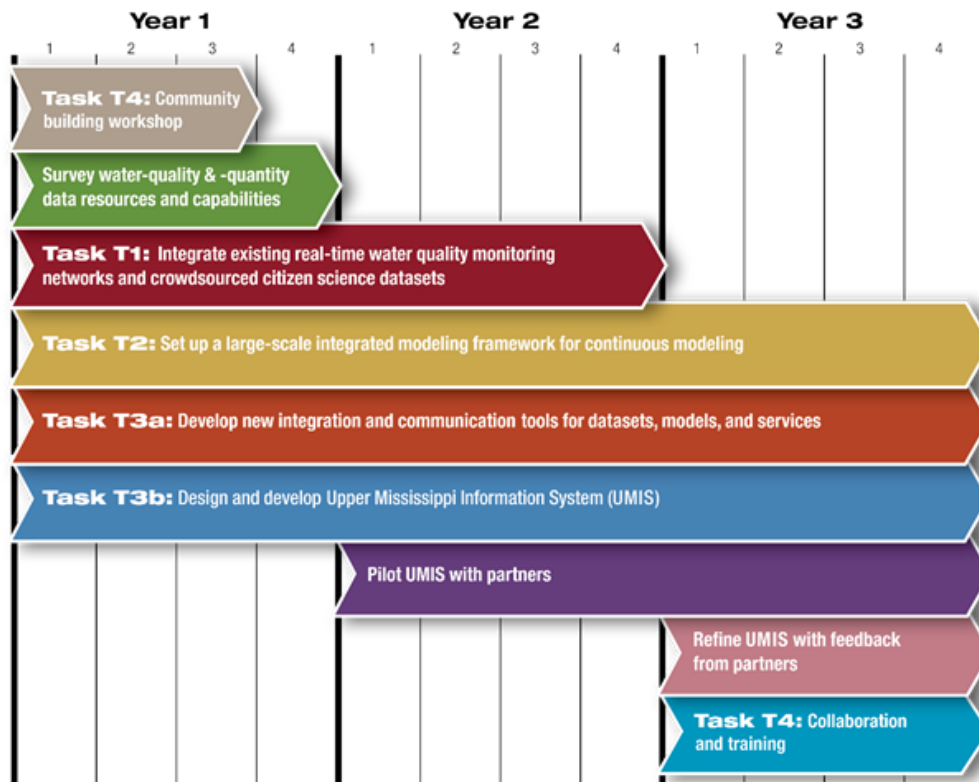


Figure 4. Project timeline with corresponding research tasks

Project Deliverables

The tangible outcomes from the project will include an integrated, modular, and scalable data-driven cyberinfrastructure platform, Upper Mississippi Information System - UMIS, with data integration, analysis, and visualization capabilities. Data integration capabilities will be used to connect data sources to the UMIS and GLTG systems. The platforms will enable researchers and stakeholders to effectively address and solve the many complex problems in water quality facing science and society through access to available data resources. The integrated and generalized structure of the platform will allow adapting the cyberinfrastructure for other geographical regions. The project will make significant interdisciplinary and synergistic impacts in the fields of water quality, digital agriculture, information systems, and big data analysis. The deliverables of the project as outlined in the timeline (Figure 4) can be listed as follows; (a) data acquisition framework for real-time and ad-hoc data streams, (b) developing specifications for data and semantics, (c) developing API specification for data access and query catalog, (d) setting up a large scale continuous modeling framework, (e) defining a model specification for model repository, (f) creating a model repository, (g) developing a client library or tools, (h) extending scale and capabilities for existing cyber systems to develop Upper Mississippi Information System, and (i) developing new visualization and communication tools.

Roles and Responsibilities

Three funded partners comprise the core implementation team, and other collaborators represent; (a) partners committed to sharing data, and (b) partners committed to attending the year one workshop to provide input on the system and to explore their interest in participation

- University of Iowa: The team has expertise in water quality monitoring, water quantity (flooding and drought), cyberinfrastructure development and operation, sensor development and deployment, computational modeling, data visualization, public outreach and communication, and evaluation. The UI team will lead the following project tasks: The year one partner workshop, project evaluation, development of UMIS, model specification and repository framework, scientific visualization tools, and data survey.
- University of Illinois Urbana-Champaign: The team has expertise in software engineering, cyberinfrastructure development, geospatial data management, water quality data and related policy. The Illinois team will lead the development of data specification, data acquisition framework, web services, and integration libraries and tools.
- Iowa State University: The Iowa State University team led by Dr. Phil Gassman at the Center for Agricultural and Rural Development has expertise in development and application of the Soil Water Assessment Tool (SWAT) watershed model to assess water and nutrient fluxes throughout North America. His team has developed SWAT models to evaluate nutrient loads for the entire Mississippi River Basin (Rabotyagov et al., 2014a). Dr. Gassman will lead the effort to evaluate monthly nutrient loading patterns in HUC12 watersheds in the UMRB for the project.
- Partners committed to sharing data: S. Geological Survey, Iowa Department of Natural Resources, U.S. Environmental Protection Agency, Illinois Nutrient Monitoring Council, Illinois Environmental Protection Agency, U.S. Army Corps of Engineers
- Partners attending the year one workshop and/or providing UMIS input: S. Environmental Protection Agency - Hypoxia Task Force, Department of the Army - Corps of Engineers, Universities Allied for Water Research (CUAHSI), Illinois Environmental Protection Agency, Iowa Soybean Association, Minnesota Agricultural Water Resources Center, Discovery Farms (WI-MN) (a component of the Minnesota Agricultural Water Resources Center), Practical Farmers of Iowa, Iowa Department of Natural Resources.

Team Background: Relevant Projects and Collaborator Synergies

This project is only possible based on several major projects and synergies that establish the feasibility of the proposed system at smaller scales and many important partnerships. Example projects this platform will build upon include:

- Iowa Water Quality Information System (Roy J. Carver Charitable Trust, \$453,264, 2015-2018; PI/Co-PIs: Weber, Demir). This supports the core development of the IWQIS as an interactive tool to improve communication and dissemination of real-time water-quality information to a wide range of stakeholders.
- Iowa Watershed Approach for Urban and Rural Resilience (U.S. Housing and Urban Development; \$12,028,551; 2016-2021; PI/Co-PIs: Weber, Demir, Krajewski). This project leverages existing IIHR resources, including the Iowa Flood Center, Iowa Flood Information System, and IWQIS to support watershed projects that decrease flooding, improve water quality, and increase community resilience.
- EAGER: Validation and Utilization of a New Tool for Citizen-Led Water Quality Monitoring in Agricultural Watersheds (NSF; \$88,714; 2017-2018; PI/Co-PI: Jones, Schilling, Demir). This project will validate a new smartphone app for testing and transmitting water quality data. (The new tool has the potential to drastically increase water quality data across the Midwest.)

This project also builds upon existing proposal partner synergies, and especially on the experience of the University of Iowa (UI) in their establishment of the Iowa Flood Information System (IFIS) and the Iowa Water Quality Information System (IWQIS), and the GLTG. The development and continued management of these systems require ongoing close collaborations with a host of partners who not only share data with the project partners, but have shared input on the systems' design and rely on the systems to serve their respective missions. One of the closest partners, for example, is the National Weather Service (NWS). The NWS provides UI with direct access to raw data from all seven NEXRAD radar that cover Iowa. Precipitation estimates from the NEXRAD data as shown on IFIS are calculated using UI's own algorithms. The NWS has direct access to data transmitted from UI's nearly 250 stream stage sensors distributed across the state. They use this data to improve and validate their flood warnings. Other examples of UI's success in the development of strong collaborations as related to water quantity and quality data include:

- S. Geological Survey (USGS) shares direct access/transmission of data from their water quality sensors distributed across Iowa with UI. That data is shared on the IWQIS visualization platform within the hydrologic context in real time.
- The Iowa Department of Natural Resources (IDNR) has contributed \$525K to UI between 2010 and 2013 for the addition of 150 stream stage sensors across Iowa. The IDNR relies solely on UI's IFIS for free access to this data for their use (e.g. monitoring droughts, managing floodplains, monitoring recreational waterways) as well as recognizes the importance of this data for public consumption.
- UI is currently leading 14 other partners across the state (IDNR, Iowa Department of Agriculture and Land Management, Iowa State University, University of Northern Iowa, Homeland Security and Emergency Management, Iowa Economic Development Authority, and eight separate community-led watershed management authorities) in fulfillment of the rural watershed components of a nearly \$100M grant to the state of Iowa by U.S. Housing and Urban Development. This project relies heavily on access to data in IFIS and IWQIS for selection, implementation, and monitoring of practices and structures to reduce flooding and improve water quality. Project partners, including citizens representing the watershed management authorities, are also helping to select placement of additional project-funded water quality sensors in their watersheds.

Collaboration with Host BD Hub and other BD Spokes

The proposed platform addresses big data needs of State's Nutrient Reduction Strategies (e.g., Iowa, Illinois, Wisconsin, Minnesota) through access to real-time (e.g. IIHR, U.S. Geological Survey, U.S. Department of Agriculture, U.S. Army Corps of Engineers, etc.) and historical (e.g. EPA STORET) data streams, which are all currently managed independently in a variety of formats. This will require close collaboration and coordination among these agencies and our participating partner institutions. As it matures, the proposed platform design will also have the flexibility to accommodate data from hitherto unidentified data and data sources impacting the UMRB. We expect our partnership with the Midwest BD Hub and its associated priority areas to further our ability to reach and partner with other entities in the Mississippi River Basin.

The proposed platform will integrate existing data and resources from existing MBDH projects and groups, and provide new capabilities and tools to support current MBDH Spokes and activities. Existing MBDH activities relevant to the project include Digital Agriculture and FEW (Food, Energy, and Water) Spokes, Networked Resilience planning project, and data science and education groups. We will utilize the following services and capabilities of the MBDH as (a) community connections/informal expertise, (b) assistance with identifying partners (e.g., companies, universities, government agencies, non-profit organizations), (b) assistance with identifying resources (e.g., data sets, data storage/compute, technical expertise, other), (c) outreach or dissemination of information about your research, events and collaboration activities, (d) facilitate data deposit or data sharing, and (e) co-hosting workshops/events.

Evaluation Plan

The University of Iowa Center for Evaluation and Assessment (CEA) will conduct the program evaluation of the project. The CEA is also the evaluator for the HUD-funded Iowa Watershed Approach for Urban and Rural Resilience, the NSF NRT, and the NSF-sponsored EAGER project. The CEA is a State of Iowa Regents-approved center that has conducted formative and summative program evaluations of more than 150 educational and social programs funded by diverse sponsors, including NSF, NIH, and the U.S. Department of Education.

Evaluation will focus on four main project components:

- The development of the community of users: The CEA will gather input from prospective users (following the year one workshop) to understand their commitment to providing data and their interest in use of the platform. (Year 1, follow-up in Years 2-3)
- The integration of the project within the culture of the Midwest Big Data Hub: The CEA will document the extent to which the project strengthens the Hub and the relevant spokes. (Years 1-3)
- Educational components: The CEA will evaluate the ways in which the project's innovations and tools are integrated into coursework at partner universities, the WSI interdisciplinary initiative, and into the NSF (NRT) program (both at the University of Iowa). (Year 3)
- Platform usage reporting: In collaboration with co-PI Demir, the CEA will assist with the synthesis and reporting of analytics related to platform use. Analytics will quantify the success of the project by using monitoring and analysis of the analytics associated with the platform usage. Reports will include number of users accessing the platform (new/returning), components and tools of interest, geolocation, demographics, time spent on the platform and components, and processed/analyzed data resources, and whether they are from rural versus urban areas. co-PI Demir will also examine user access in response to specific weather events (e.g., long period of rainfall). The CEA will gather input from users as to the purpose for and satisfaction with their interaction with the platform for decision-making. (Year 3)

Long-Term Sustainability

This proposal represents an investment in a very sustainable tool that is expected to continue to grow even after the conclusion of NSF funding. This is demonstrated by the current systems' (IWWIS and GLTG) success, as well as the public's increasing desire for mitigation of water-quality issues in the Mississippi River Basin and across the nation. The recent landmark lawsuit filed by the Des Moines Water Works against several drainage districts in central Iowa drew national attention to the water-quality issues that have plagued the Midwest for decades. Despite the outcome of the lawsuit, IIHR's experience working in the watersheds across the state through the Iowa Watershed Approach (IWA) has revealed that stakeholders – from producers to policy-makers – are eager for information and solutions. Iowa watershed groups rely heavily on data from IIHR's water-quality sensors and models to help identify the best projects (ponds, wetlands, retention basins, bioreactors, etc.) and the most effective project sites to improve water quality in their basins. Based on our experience, the IWA will likely attract more interested landowners than this nearly \$100M project can support, even with a required cost share from landowners. As projects like the IWA continue to gain in popularity, the desire for easily accessible information will likewise grow.

As stated previously, states have made accountability part of the process for use of public funds by clearly stating the goals in their state nutrient reduction strategies. The best way to accurately quantify water-quality improvements is through the collection of publicly accessible data to directly measure the improvements or model the expected outcomes. This means the states in the UMRB will also have an interest in the longevity of UMIS.

Throughout the project, we will establish a long-term sustainability plan with support from the Midwest Big Data Hub, as well as project partners and collaborators. For example, both the U.S. Army Corps of Engineers and the U.S. Environmental Protection Agency have committed to helping to identify resources to sustain the UMIS. Other potential resources will be sought from federal sources, (e.g., U.S. Department of Agriculture, U.S. Environmental Protection Agency), state agencies (e.g., natural resources departments) and direct state appropriations, private foundations (e.g., McKnight Foundation, Walton Foundation), NGOs, and other entities directly affected by or tasked with addressing water-quality issues at local, state, regional, and/or national levels. The project team will also leverage partners' scientific cloud capabilities to host the system in the future. The researchers involved in the proposed project already have a successful track record of funding from a variety of sponsors in support of the existing platforms.

Education

The project's research innovations and cyber tools may be integrated into courses at the partner universities and serve as a "playground" for educational events such as student hackathons. Students will receive training for careers in water quality, digital agriculture, information systems, and big data analysis. The resulting tools will be available to staff at relevant agencies, NGOs, and other organizations through webinars, workshops, and other professional development opportunities.

We will also be able to leverage the existing NSF Research Traineeship (NRT) grant awarded last year to the University of Iowa, and housed in the Department of Civil and Environmental Engineering and IIHR-Hydroscience & Engineering. The goal of this five-year grant is to develop a new graduate program on Sustainable Water Development to address the challenges at the intersection of the food, energy and water systems (FEWs). An important component of this program is represented by the students' interactions with both rural and urban communities in Iowa. The proposed framework will provide the tools and information to present data-driven evidence to support statements about the water quality conditions of the communities that the will interact with.

Furthermore, the Water Sustainability Initiative (WSI) at the University of Iowa is within IIHR-Hydrosience & Engineering. WSI represents an interdisciplinary team of faculty members from different departments across campus; their goal is to develop solutions needed for a sustainable management of the water resources and to educate future generations of citizens: the proposed system will provide them with the information required to develop robust strategies to address water quality issues, and a platform to communicate the state of the health of our rivers to the communities at large.

The research innovations and cyber tools emerging from the project will be integrated into various courses at UI, UIUC, UW and other collaborating universities. The proposed platform, tools and services would serve as an excellent resource for various class activities in the courses taught by the PIs, such as project-based big data analysis and visualization. The proposed research will utilize resources of many research centers and institutions. All these centers will offer extensive research and educational opportunities related to the proposed work at undergraduate and graduate levels. The proposed platform will be an extensive resource for Midwest Big Data Hub and its activities related to the "Digital Agriculture" and "FEW" spokes, and be integrated to regional and local community events (e.g. All Hand Meeting) of the MBDH. PI Demir organizes a Midwest Big Data Hackathon at the University of Iowa as an annual event for bringing students together in Big Data domain. The platform will be a critical playground for the activities at that hackathon and other similar events in the Midwest.

Broader Impacts of the Proposed Work

This project will improve measurement and management of water resources in the Upper Mississippi River Basin (UMRB) through a cyberinfrastructure framework that will facilitate large-scale data integration, analyses, and visualization using data-enabled information technologies. This project ties directly to three National Academy of Engineering Grand Challenges as we will engineer the tools of scientific discovery with a platform that allows comprehensive integration of water quality data in the UMRB to guide research and land management practices to better manage the nitrogen cycle, which will lead to improved access to clean water. The immediate impacts of this project include facilitating a centralized platform for data access, integration, and scientific discovery that leads directly to improved measurement and management of water resources. The long-term implications are the broader scientific community's potential adoption of data-enabled cyberinfrastructure tools and services through the Midwest Big Data Hub to expand the reach to the entire Mississippi River, and later the nation. The proposed project will allow the development of new and strengthening of existing collaborations around water-quality big data among researchers at the University of Iowa, University of Illinois, Iowa State University, and National Great Rivers Research and Education Center. Development of the platform also necessitates direct engagement with many state and federal agencies, watershed managers, and NGOs. All of these groups, and the general public, will have access to the platform to better monitor, understand, and characterize water quality problems at local and regional watershed scales.

Results from Prior NSF Support

Krajewski: NSF# 1464613I \$161,287; 03/01/2016 to 02/28/2018; *Acquisition of Equipment for Calibration and Engineering Tests for University of Iowa's XPOL Radars*. This project is for acquisition of equipment only. The intellectual merit of this project is the higher quality data that the University of Iowa's XPOL radar network will collect in support of hydrologic and precipitation research. The broader impacts include the many field experiments, research projects, and student educational opportunities that the radar network will support. There are no results or publications to date from this recently-awarded project as the calibration equipment is on order.

Demir: NSF# 1114978; \$1,211,485; 09/01/2011 to 02/29/2018; *CNH: People, Water, and Climate: Adaptation and Resilience in Agricultural Watersheds*. This project is investigating how coupled natural and human systems respond to changes in climate, economics, and policies that operate over large geographic and time scales. The intellectual merit is that we will develop a better understanding of how land use practices in the agricultural Midwest adapt to changes in policies, economic conditions, and climate. The broader impacts include the effect of this research on producers to adopt management practices that maximize sustainability and resilience under changing conditions. (Demir was added as senior personnel on supplemental part of the project. There are no products to report yet on the supplement.)

Jones and Schilling: NSF #1743991; \$88,714; 09/01/2017 to 08/31/2018; *EAGER: PPER: Validation and Utilization of a New Tool for Citizen-Led Water Quality Monitoring in Agricultural Watersheds*. This project involves testing a new phone app that allows citizens to accurately test and transmit water quality data to a central platform. The intellectual merit is the project builds on existing IIHR expertise, data platforms, and partnerships. The broader impacts include the potential adoption of a new tool to test water quality, and the resulting availability of more water-quality data to support researchers, agencies, and producers. There are no accomplishments or publications to report as this is a new project.

Jong Sung Lee: NSF# ACI-1261582; \$10,519,716; 10/01/2013 to 09/30/2018; *CIF21 DIBBs: Brown Dog (BD)*. BD aims to develop two distributed and extensible open source services towards aiding scientific reproducibility and reuse, the Data Access Proxy for file format conversions, and the Data Tiling Service for metadata extraction and indexing, removing these common and heterogeneous data wrangling steps from contemporary scientific workflows. The intellectual merit is that the services will allow third-party tools and applications to move closer towards being agnostic to file formats and uncured data collections being more readily leveraged and address a need for broad extensibility, heterogeneous usage, provenance/validation/information loss tracking, support for large datasets and computationally intensive analysis. The broader impacts are that the services address needs within the ecology, civil & environmental engineering, and social science communities as well as geoscience, biology, agriculture, material science, genomics, medicine, and the humanities. The project addresses a general public use case towards its own long-term sustainability, again filling a role that a broad base of users could utilize. BD is in a pre-beta development phase with a beta release scheduled for early 2017. Products include seven publications (Padhy et al., 2015a,b; 2016a,b; Jansen et al., 2016; McHenry et al., 2014, 2016), six user cases (Dietze et al., 2015; Woo et al., 2014, 2015a,b; Rabotyagov et al., 2013, 2014a-b), 2015; Rai, 2016; Slavenas et al., 2016;) and one tutorial.

Gassman: NSF# DEB-1010259; \$655,289; 10/1/10 to 9/30/14; *CNH: Collaborative Research: Northern Gulf of Mexico Hypoxia and Land Use in the Watershed: Feedback and Scale Interactions*. Hypoxia is a growing environmental problem for coastal habitats in the United States and many other countries. Each year an oxygen-depleted (hypoxia) zone forms in the Northern Gulf of Mexico, negatively impacting coast and marine fisheries and local economies. The intellectual merit is that the research produced the first complete modeling system that traced agricultural land-use decisions, made at the field scale in the Upper Mississippi, Ohio, and Tennessee Basins, through both environmental and hydrologic components, to downstream water quality effects, including the size of the hypoxic zone in the Gulf of Mexico. The broader impacts are that the project demonstrated how to integrate human and natural process models using the powerful tools of evolutionary algorithms. Furthermore, this project strengthened an interdisciplinary collaboration among scientists at five different institutions, and supported the training of 5 graduate and 2 undergraduate students. Nine refereed journal articles were published (Jha and Gassman, 2014; Jha et al., 2015; Kling et al., 2014; Panagopoulos et al., 2014, 2015a,b; Rabotyagov et al., 2013, 2014a-b).

Weber: None

Warner: None

