

City University of New York (CUNY)

## CUNY Academic Works

---

International Conference on Hydroinformatics

---

2014

### Web-Based Engine For Discovery Of Observations Using Landscape Units

Haowen Xu

Abbas Ali Amir

Marian Muste

[How does access to this work benefit you? Let us know!](#)

More information about this work at: [https://academicworks.cuny.edu/cc\\_conf\\_hic/182](https://academicworks.cuny.edu/cc_conf_hic/182)

Discover additional works at: <https://academicworks.cuny.edu>

---

This work is made publicly available by the City University of New York (CUNY).  
Contact: [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu)

## **WEB-BASED ENGINE FOR DISCOVERY OF OBSERVATIONS USING LANDSCAPE UNITS**

XAOWEN XU (1), ABBAS ALI AMIR (1), MARIAN MUSTE (1)

(1): *IIHR-Hydroscience & Engineering, The University of Iowa, Iowa City, IA 52246, USA*

The web-based engine presented in this paper facilitates the customized discovery, retrieval, and visualization of hydrologic and water quality data available in Iowa watersheds irrespective of watershed size and data provenance (citizen science, academic or management settings). The search engine locates and displays datasets in the upstream drainage area from a point of interest specified by the user on the map. The engine is built atop of the Arc GIS data model and in conjunction with the National Hydrography Dataset Plus (NHD Plus) geo-spatial dataset.

### **Introduction**

Investigations of coupled human-natural processes related to the water cycle are best studied using a commensurate landscape unit that matches the spatial extent of the process evolution and its interactions with the surroundings. The most often used landscape units used in the hydrology/hydraulics domain are those represented by catchments, sub-basins, basins, watersheds. Most often, however, the landscape units are expressed in terms of hydrologic unit codes (HUC) as defined by the US geological Survey [1]. Natural processes are best characterized as continuous flows of material and energy, while humans manage the world as discrete units in space and time (e.g., reporting periods, growing seasons, jurisdictions, farm fields). Consequently, the capability to efficiently delineate the extent of the watershed that affects the water fate at any location on a map along with the data on the main hydrological, bio-geo-chemical and human characteristics within the watershed are essential for investigating observations in natural environment. The search engine discussed in the present paper is designed to identify various observations acquired in the upstream drainage area from a point of interest specified by the user on an interactive map. The point can be selected on or outside the stream network using a web mapping interface. Most of the data under consideration here is directly accessed from the server where the original data is stored and curated.

The data and attributes associated with the variables (e.g., stream flow, water quality, weather, etc) are stored in the geodatabase supporting the application. The base map for the search engine is the National Hydrography Dataset Plus V2.0 (NHD Plus) and Geometric Network analysis are applied to develop the model on the Arc GIS platform. The user input is defined as the point of interest for the search. Subsequently, the drainage area upstream from the point of interest is identified and visualized using mapping functions available in the NHD Plus library. Ancillary information provided by the NHD database and other relevant attributes of the data for the discovered point of observations are also provided.

### **Host server and existing cyberinfrastructure**

The search engine is built a top of several cyberinfrastructure components assembled by our team or transferred from other resources. In essence, the application presented here is based on the National Hydrography Dataset (NHD), Arc GIS, and the Iowa Watershed Data & Information System. They are briefly described below. NHD is a digital vector geospatial database on surface water hydrography obtained from topographic maps and additional sources by a partnership among the U.S Geological Survey (USGS) and more than 50 government agencies at federal, state and local levels. In 2002, the NHD was available nationwide as medium resolution at 1:100,000-scale. By the end of 2006, a new high-resolution NHD at 1:24,000-scale was complete with national wide coverage [2].

Besides the basic surface-water geometry including river flow lines which is a network dataset, the NHD offers many other features and water related intelligence that have made it the standard hydrography coverage for the nation. The NHD river flow lines featured with flow-direction and linear referencing gives the dataset more GIS capability which allows users to trace downstream or upstream inside the surface water network and locate hydrographic infrastructure along the path. Stream gages, water bodies and fish habitats are also included in the dataset as events [2]. In the NHD plus Version 2.0, Nation elevation dataset (NED), Watershed Boundary Dataset (WBD), which defines the areal extent of surface water drainage to a point, accounting for all land and surface areas and river catchment characteristics are also incorporated into the NHD dataset [3]. With comprehensive data storage capability and advanced GIS functionality provided by various features, NHD becomes a perfect tool in GIS-based watershed management

To design the application algorithm, Arc GIS is used as a development platform [4]. The Geographic Information System (GIS) is a digital system designed to collect, record, manage, analyze, and visualize all types of geographical data and geographically related features or events; for instance, watersheds or pollution spread in a river. Embedded with many powerful toolset, the Arc GIS is capable of performing analysis on both raster file and vector file. The search engine is assembled with the GIS model builder, where all the existing geo-processing tools can be arranged in logical sequence with optimum settings to produce a reasonable result [5]. The resulting model from the GIS model builder can be exported as both Geo-processing service and Python scripts, as illustrated in [6]

The search engine described herein is developed for the Iowa Watershed Data & Information System, IOWADIS [7]. The geo-platform has been developed by The University of Iowa interdisciplinary research teams since 2003 with support from the National Science Foundation [8]. These successive projects enabled researchers to assemble a suite of digital environmental observatories associated with watersheds (or jurisdictions) of widely different areas [9]. The common denominator of these platforms is the assemblage of off-the-shelf components developed by Consortium of Universities for the Advancement of Hydrologic Science, Inc.-Hydrologic Information System (CUAHSI-HIS) project in customized workflows serving various purposes [10, 11]. Specifically, the digital observatories entail the following CUAHSI-HIS project components:

- Observational Data Model (ODM) - a relational database schema
- Streaming Data Loader (SDL) – a file based data loader scheduled to continuously update datastreams
- WaterOneFlow (WOF) – SOAP-based web services customized for hydrologic data using WaterML information exchange schema
- Time Series Analyst – software package for time series visualization

These components are freely available and documented with comprehensive manuals and other publications provided on the CUAHSI-HIS project website [12]. The state-wide data and information system is built on ArcGIS Server 10.2.1 taking advantage of ESRI's streamlined client-side Javascript API. In essence, the IOWADIS interactive window contains four blocks, as illustrated in Figure 1:

- a) Data layer activation (time series, shape files) – this block provides access to several data layers (one at a time or concurrently) that can be mapped through their location on the interactive map.
- b) Search action commands – These are action buttons that are used for the resent application (they will be described I detail next)
- c) Interactive map – the map area covering the entire state of Iowa
- d) Navigation tools – tools allowing users to select customized map visualizations (zoom out, zoom in or reset the map its original condition).

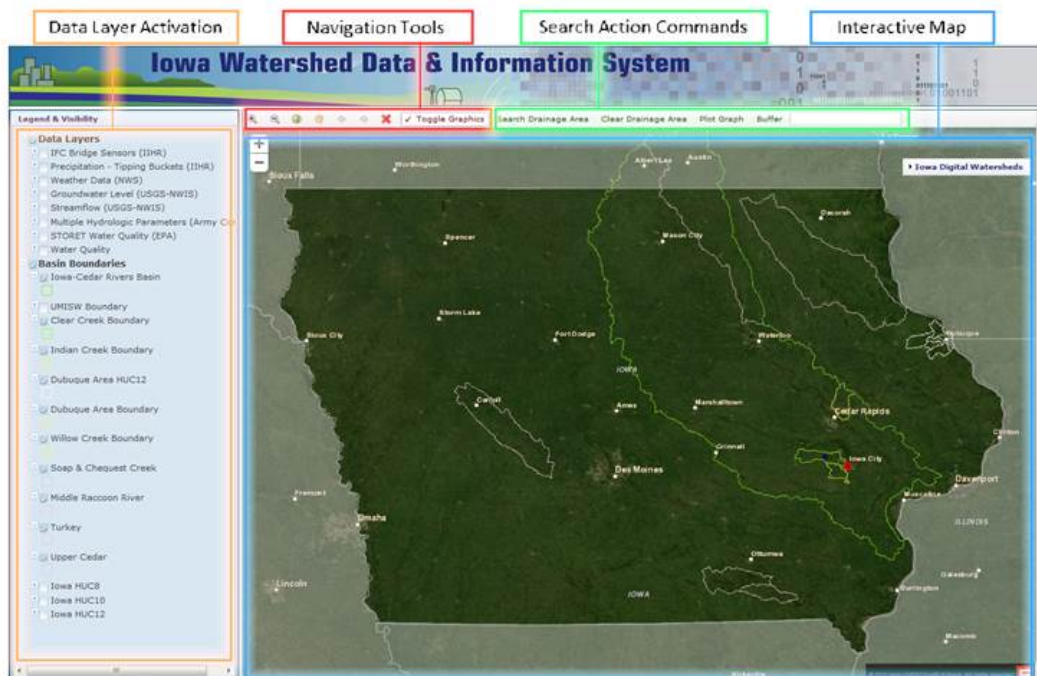


Figure 1. Iowa Watershed Data & Information System

### Search engine back-end

The search engine is developed using the Model Builder platform and Python scripting. During the execution there are two stages and several sequential steps carried out. They are described briefly below:

#### Stage 1 Geo-processing Model

1. Load the NHD 2.0 dataset and related watershed vector files
2. Set the point dynamically set by the user on the map. Similar to the route search on Google Maps, a user can select a location interactively by dropping a pin on map. The input location is Geo-referenced and combined with the NHD plus version 2.0. This point is referred to herein as POI.

3. Based on the user input, the model identifies the river segment whose catchment area contains the point of interest (POI).
4. Once the river reach is identified, the model traces all the upstream paths of the selected river segments. Flow directions for each river segment are defined using the NHD river networks database.
5. The model identifies all sub-watersheds corresponding to the reaches of the river network located upstream from the POI location. NHD Plus also identifies stream gauges located within the drainage area and displays their locations onto the map.

*Stage 2 Publish the Model on the server as an application.*

1. Publish the model on the GIS server using GIS server manager, so the model becomes a web service.
2. Connect the web service to the Iowa Watershed Data & Information interface

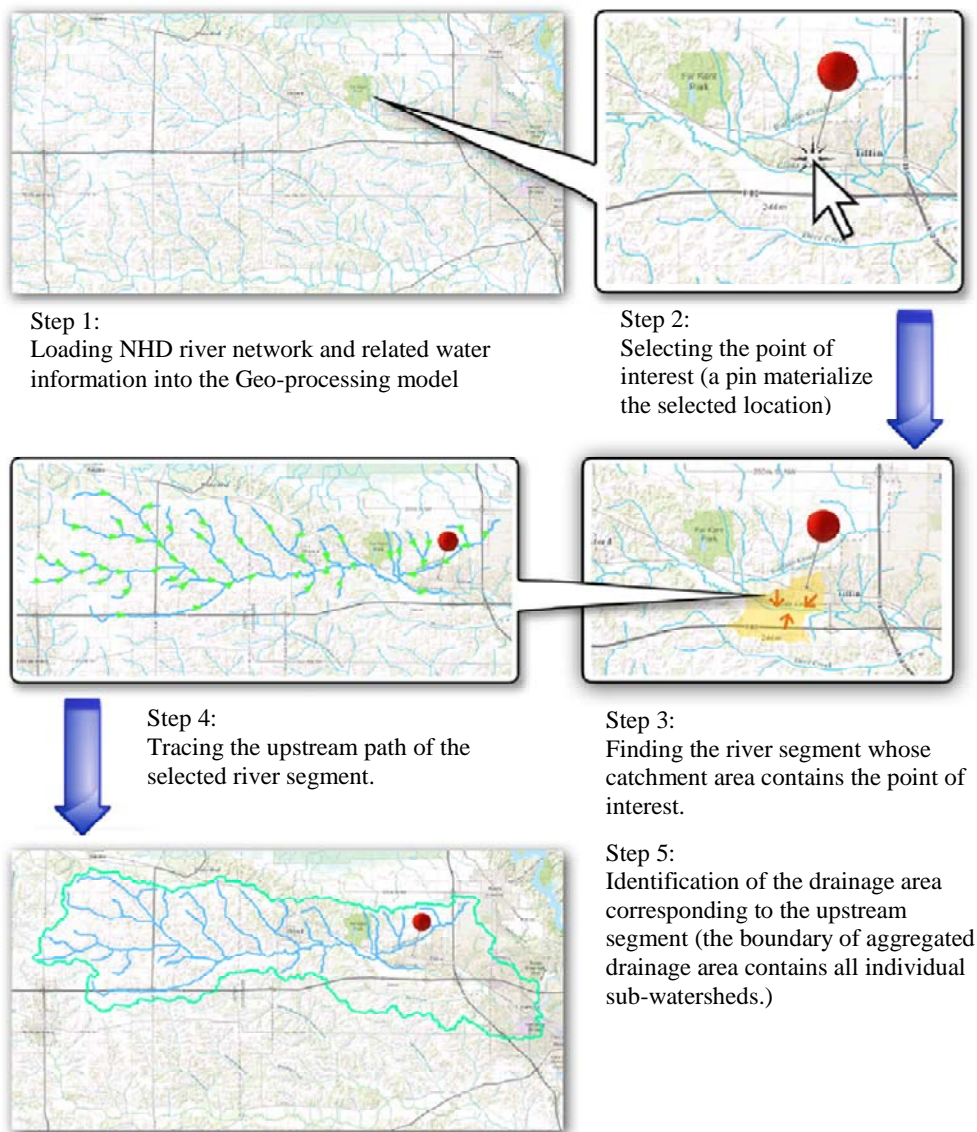


Figure 2. Flow diagram of the model

### **Search Engine Front-End**

In the Iowa Watershed Data & Information System, the search engine is activated by the commands grouped in the Search Action Area illustrated in Figure 1 and 3. The following actions are triggered when the user select one of the following buttons :

1. Search Drainage Area (illustrated as step 1 in Figures 3.a and 3.b) – Clicking this button triggers the search for the drainage area search upstream POI (a yellow pin on the map).
2. Clear Drainage Search – This button clears all the drainage area search results and resets the map to its original condition.
3. Plot StreamFlow Graph – This button provides visualization of a time-series data for the selected layer. This functionality is still under progress and not discussed here
4. Buffer (illustrated as step 3 in Figure 3.b) – If the drainage area search doesn't result in having any data point inside the polygon (as discussed below), user has an option to manually extend the search by keying in a buffer area in miles.

Clicking on a point on the interactive map and subsequently on the 'Search Drainage Area' button triggers a request to the ArcGIS server. REST based services in the ArcGIS server return the search result in the form of a polygon s illustrated in Figure 3.a and 3.b by step 1. The next step in the data discovery is selecting the data of interest from the available data layers so that data points in the polygon can be visualized, retrieved and/or analyzed. For this purpose the user has to select one of the data layers (i.e., streamflow) as illustrated by step 2 in Figures 3.a and 3.b. The follow up action is that the data points within the identified polygon are displayed (blue dots in Figure 3.a).

If there are no data points found within the drainage area associated with the POI, an extended search is automatically triggered in a buffer area surrounding the initial polygon. The buffer is shown in Figure 3.b as a dark blue contour map around the identified polygon. The default thickness of the buffer is 5 miles outside the polygon perimeter. The buffer is automatically increased for maximum of four times with 5 miles increments until at least one data point is found. Once the data is found in the surrounding region they are displayed as blue dots on the map. If there is no data point found even after four rounds of automatic extended search, there is an option for the user to enter the buffer value manually. The buffer value (in miles) when entered by the user and the 'Buffer' button clicked, data points are yet again searched for. Similar to the automatic extended search behavior, if any data point is found within the buffered region, it is highlighted in blue color. In order to repeat the search for a different POI, the user selects first 'Clear Drainage Search' button which will clear and reset the map.



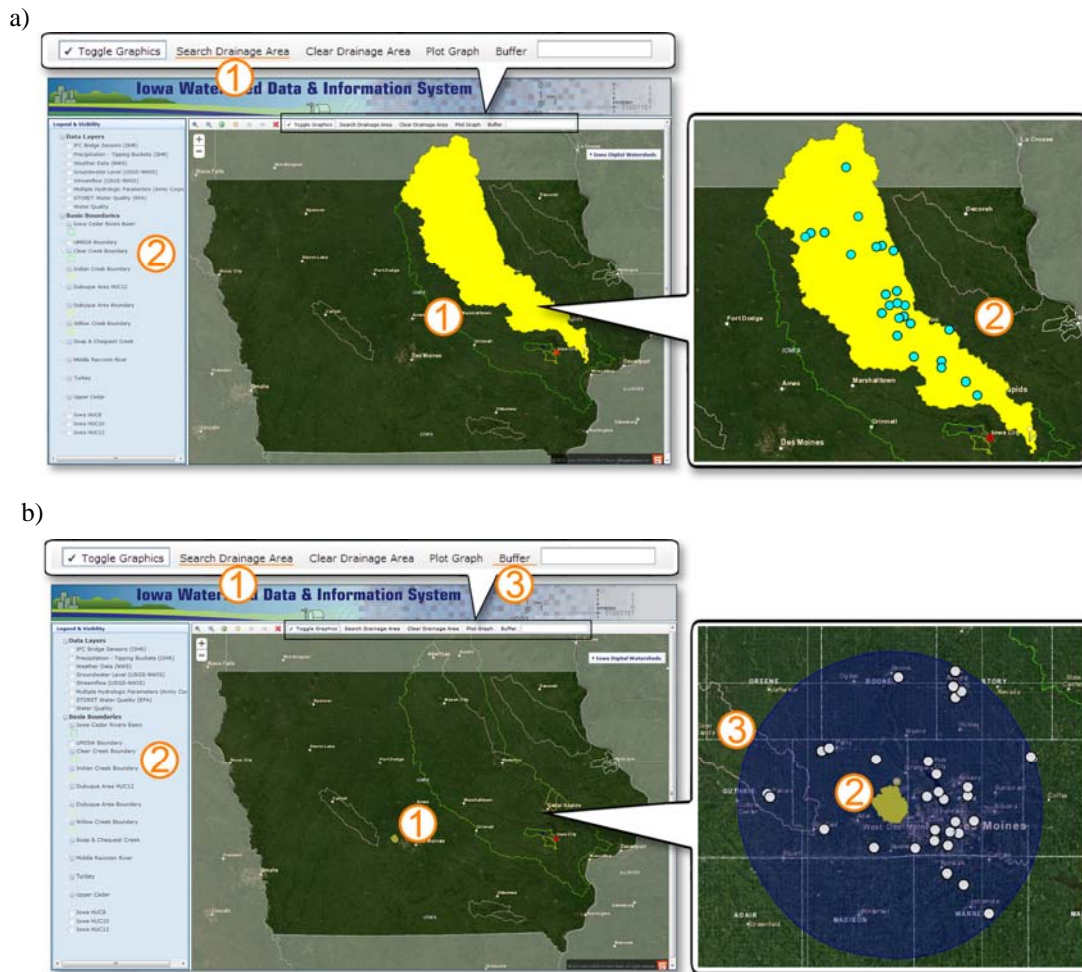


Figure 3. Search workflows for data discovery: a) search resulting with a drainage area with data within; b) search resulting with a drainage area without data. Subsequent search can be requested until data points data are found using user-specified distances from the watershed.

The search can be simultaneously applied to multiple data layers using the above-described protocol. Following the discovery of the data points in the search area, the web-engine can be used to visualize the data. Selection of any discovered data point leads to a small popup information window as shown in Figure 4. This window displays the variable name, station ID, data provider, time series extent, and the time stamp for the last recorded value. This window is linked to the HydroServer Time Series Analyst (TSA) developed by the CUAHSI-HIS that provides several choices for data visualization [13]. The visualization is triggered by choosing the ‘Stats & Graphs’ link. Subsequently, a window with multiple options for visualization of the data opens up as shown at the bottom of Figure 4. Besides the direct plot of the time series for selected time intervals, the TSA plots on the fly the probability, histogram, or box and Whisker plots for the time series displayed. The summary statistics is also provided as numerical values for each of the visualization choices. The data associated with the plots can be seen as an HTML table or as a comma separated values files when asking for ‘Export myDB’ button. The metadata is exported as XML file using the ‘Export Metadata, button.

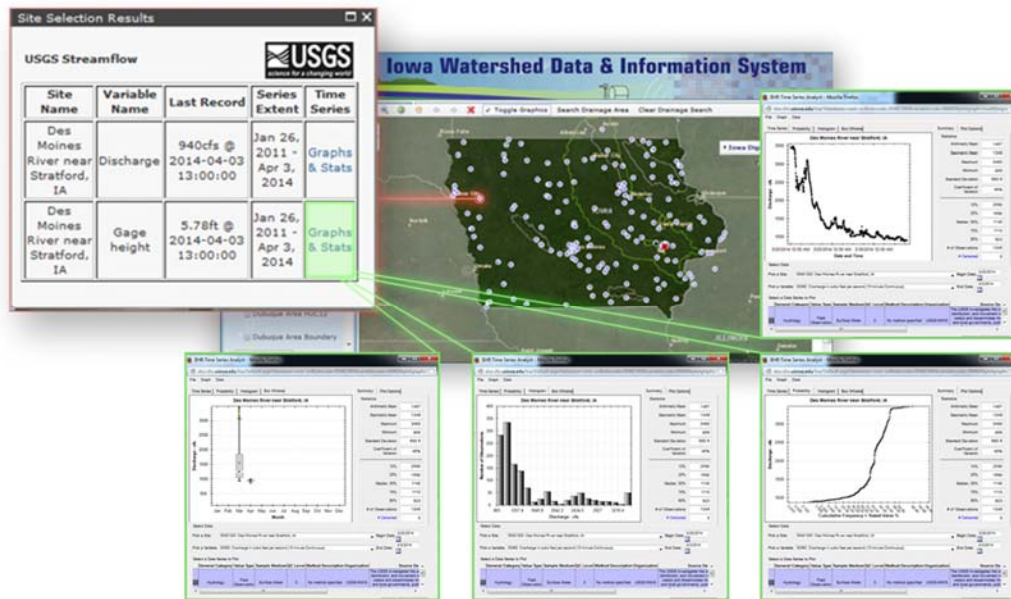


Figure 4. Data visualization alternatives for individual data points of the data layers in the Iowa Watershed Data & information System (iowadis.org)

### Lesson Learned

The search engine presented in this paper illustrates an efficient tool for facilitating documentation of various aspects of water quantity and quality at a point of interest in the watershed. It does so by tracing the available water-related observations upstream from a specified point to answer the question: from where the water comes to this point? The search is conducted over data and information from distributed sources using navigation tools and layouts that allow making useful connections with other observations in the drainage area upstream from a specified point. The application holds promise to benefit users from watershed management and watershed resources researcher communities irrespective of their location as the customized system is built on a database that has nation-wide coverage (NHDPlus).

The paper presented herein illustrates that by using a combination of off-the-shelf and community-developed software (i.e., Arc GIS and CUAHSI-HIS components) it is possible to assemble customized applications for discovery and visualization of the data that fit various local research and water resources research needs. The approach taken by this team removes some of the burden in handling big volume of data from disparate resources leaving more room for assembling them for further analysis and synthesis using efficient and intuitive tools that were not available so far. While the customization of various software is not a particularly difficult task for a computer science researcher/student, it represents a considerable hurdle for a domain specialist with conventional computer/computational background. This first-hand observation suggests that future research at natural scales will necessarily include a cyberinfrastructure specialist as an integral part of the domain researcher group for the design, implement, and use these types of web-engines that considerably increase the productivity and efficiency of the collaborative actions.



## Acknowledgements

This work is funded in part by the National Science Foundation Awards through the Coupled Natural-Human Systems (award# 1114978) and Critical Zone Observatories (award# 1331906). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## REFERENCES

- [1] Seabar R. P., Capanos F.P. and Knapp G.L. "Hydrologic unit maps," United States Geological Survey Water-Supply Papers, 2294 (1987), pp. 1-63.
- [2] Simley J., "The National Hydrography Dataset leverage the data," Geoworld, (2006), pp. 30-31.
- [3] McKay L., Bondelid T., Dewald T. et al., "NHDPlus Version 2: User Guide," Horizon Systems, Herndon, VA.
- [4] ESRI, "ArcGIS Desktop: Release 10.2.1," Environmental Systems Research Institute Redlands, CA, (2011).
- [5] Xu H., "Search engine for Iowa Watershed Information System," Honors Thesis Program, University of Iowa, Iowa City, IA, USA, (2013).
- [6] ESRI, "ArcGIS API for JavaScript Environmental Systems Research Institute Redlands, CA, (2011).
- [7] Iowa Watershed Data and Information System ([iowadis.org](http://iowadis.org)), IIHR-Hydrosience & Engineering, Watershed, The University of Iowa, Iowa City, IA, USA, (last accessed, April 1, 2014)
- [8] National Science Foundation, ([nsf.gov](http://nsf.gov)), Arlington, VA, USA, (last accessed, April 1, 2014)
- [9] Muste M., Chapter in "Comprehensive Water Quality and Purification," Ahuja S. (Ed), Ch. 13: "Information-centric systems for underpinning sustainable watershed resource management," vol 4, Elsevier, (2014), pp. 270-298.
- [10] Muste M., Kim D., Arnold N., Whiteaker T., Just C. and Kruger A. "Digital watershed inception using community project components," Special Issue on Cyberinfrastructure for Hydrology and Environmental Engineering, *Proceedings of Institution of Civil Engineers, Water Management*, 163(1), Thomas Telford Journal, London, UK, (2010). pp. 13-25.
- [11] Muste M., Bennett D., Secchi S., Schnoor J., Kusiak A., Arnold N., Rapolu U., Deng D. and Mishra, S. K.. "End-to-end cyberinfrastructure for decision-making support in watershed management," *Journal of Water Resources Planning and Management*, American Society of Civil Engineers, Reston, VA, 139(5), (2013), pp. 565-573.
- [12] Consortium of Universities for the Advancement of Hydrologic Science, Inc. ([cuahsi.org](http://cuahsi.org)), Medford, MA, USA, (last accessed, April 1, 2014).
- [13] HydroServer Time Series Analyst, Consortium of Universities for the Advancement of Hydrologic Science, Inc. – Hydrologic Information System ([his.cuahsi.org/hstsa.html](http://his.cuahsi.org/hstsa.html)), Medford, MA, USA (last accessed, April 2, 2014).